

# Varieties of Unranked Tree Languages

Magnus Steinby\*      Eija Jurvanen†      Antonio Cano‡

October 27, 2015

## Abstract

We study varieties that contain unranked tree languages over all alphabets. Trees are labeled with symbols from two alphabets, an unranked operator alphabet and an alphabet used for leaves only. Syntactic algebras of unranked tree languages are defined similarly as for ranked tree languages, and an unranked tree language is shown to be recognizable iff its syntactic algebra is regular, i.e., a finite unranked algebra in which the operations are defined by regular languages over its set of elements. We establish a bijective correspondence between varieties of unranked tree languages and varieties of regular algebras. For this, we develop a basic theory of unranked algebras in which algebras over all operator alphabets are considered together. Finally, we show that the natural unranked counterparts of several general varieties of ranked tree languages form varieties in our sense.

This work parallels closely the theory of general varieties of ranked tree languages and general varieties of finite algebras, but many nontrivial modifications are required. For example, principal varieties as the basic building blocks of varieties of tree languages have to be replaced by what we call quasi-principal varieties, and we devise a general scheme for defining these by certain systems of congruences.

**Keywords:** tree language; unranked tree; syntactic algebra; variety of tree languages; unranked algebra

## 1 Introduction

In its prevalent form, the theory of tree automata and tree languages (cf. [30], [12], [13] or [8] for general expositions) deals with trees in which the nodes are labelled with symbols from a ranked alphabet; in a ranked alphabet each symbol has a unique nonnegative integer rank, or arity, that specifies the number

---

\*Department of Mathematics and Statistics, University of Turku, FI-20014 Turku, Finland, email: steinby@utu.fi

†Department of Mathematics and Statistics, University of Turku, FI-20014 Turku, Finland, email: jurvanen@utu.fi

‡Departamento de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia, Camino de Vera s/n, P.O. Box: 22012, E-46020 - Valencia, email: acano@dsic.upv.es.

of children of any node labelled with that symbol. Thus a ranked alphabet may be viewed as a finite set of operation symbols in the sense of algebra, and then trees are conveniently defined as terms and finite (deterministic, bottom-up) tree automata become essentially finite algebras. As a matter of fact, the theory of tree automata arose from the interpretation of ordinary automata as finite unary algebras advocated by J.R. Büchi and J.B. Wright already around 1960 (cf. [12] or [13] for notes on this subject and references to the early literature). Universal algebra has offered the theory of tree automata a solid foundation, and the definition of trees as terms links it naturally also with term rewriting. Nothing of this would be lost even if each symbol in a ranked alphabet is allowed a fixed finite set of ranks. However, when trees are used as representations of XML documents or parses of sentences of a natural language, fixing the possible ranks of a symbol is awkward. It is in particular the study of XML that propels the current interest in unranked tree languages (cf. [6], [17], [16], [8] or [21], for example).

Actually, unranked trees are nothing new in the theory of tree languages. Let us note just two early papers, published in 1967 and 1968, respectively. In [29] Thatcher defines recognizable unranked tree languages, proves some of their basic properties, and establishes a connection between them and the derivation trees of extended context-free grammars; the paper is motivated by the study of natural languages. Recognizability is defined using “pseudoautomata”, a concept attributed to Büchi and Wright, in which state transitions are regulated by regular languages over the state set. This idea reappears in various forms in most works on recognizable unranked tree languages, and also in this paper pseudoautomata play a central role (we call them regular algebras). In [18] Pair and Quere consider hedges (that they call ramifications), i.e., finite sequences of unranked trees, and they introduce a new class of algebras, binoids, in terms of which the recognizability of hedge languages are defined and discussed. Also hedges have become a much used notion in the theory of unranked tree languages (cf. [28, 6, 5, 4], for example). However, we shall consider just unranked trees.

The varieties to be studied here contain tree languages over all unranked alphabets and leaf alphabets, and we take as our starting point the theory of general varieties of (ranked) tree languages presented in [25]. However, in addition to the modifications to be expected, some novel notions are needed. On the other hand, the formalism is actually simplified by the fact that symbols have no ranks.

The paper is organized as follows. In Section 2 we recall some general preliminaries, introduce unranked trees and some related notions. In addition to an unranked alphabet, that we call the operator alphabet, we use also a leaf alphabet for labeling leaves only. If  $\Sigma$  is an operator alphabet and  $X$  a leaf alphabet, then unranked  $\Sigma X$ -trees are defined as unranked  $\Sigma$ -terms with variables in  $X$ , and sets of such terms are called unranked  $\Sigma X$ -tree languages. This arrangement with two alphabets will be convenient for the algebraic treatment of our subject but, as we shall demonstrate, it is also natural in typical applications.

In Section 3 we develop the basic theory of unranked algebras in a way that allows us to consider together algebras over different operator alphabets. Here

we can follow quite closely the corresponding generalized theory of ordinary (i.e., ranked) algebras as presented in [25]. In the next section we consider the unranked algebras in terms of which recognizability is defined. We call them regular algebras, but they are precisely the pseudoautomata of Büchi and Wright mentioned above. We show that the class of regular algebras is closed under our generalized constructions of subalgebras, epimorphic images and finite direct products. Thus they form the greatest variety of regular algebras (VRA). By first proving a number of commutation and semi-commutation relations between the class operators corresponding to the various constructions of algebras, we derive a representation for the VRA generated by a given class of regular algebras similar to Tarski's classical HSP-theorem (cf. [7] or [3], for example). The regular congruences considered in Section 5 are intimately connected with regular algebras. Indeed, the congruences of a regular algebra are regular, and the quotient algebra of an unranked algebra with respect to a congruence is regular exactly in case the congruence is regular.

In Section 6 we introduce syntactic congruences and syntactic algebras of subsets of unranked algebras. Also here it is convenient to consider these notions on this general level. Syntactic congruences and syntactic algebras of unranked tree languages are then obtained by viewing them as subsets of term algebras. All these notions are natural adaptations of their ranked counterparts (cf. [23, 24, 26] or [1]). Our syntactic congruences of unranked tree languages appear also in [6] as 'top congruences'. In [4] the term 'syntactic algebra' designates a different notion that is associated with hedge languages. Similarly as in [25], we shall also need reduced syntactic congruences and algebras obtained by merging symbols that are equivalent with respect to the subset considered. In Section 7 we define an unranked tree language to be recognizable if it is recognized by a regular algebra. This definition is essentially the same as that of [29] and equivalent to other definitions that use finite automata. As one would expect, an unranked tree language is recognizable if and only if its syntactic algebra is regular, and the syntactic algebra is in a natural sense the least unranked algebra recognizing any given unranked tree language. We also show that the syntactic algebra of any effectively given recognizable unranked tree language can be effectively constructed; here this is less obvious than in the ranked case as the operations are infinite objects and there are infinitely many trees of any given height  $\geq 1$ .

In Section 8 we introduce varieties of unranked tree languages (VUTs). Such a variety contains languages for all operator and leaf alphabets. Similarly as in the ranked case, a VUT is usually most naturally defined in terms of congruences of term algebras, and hence we introduce also varieties of regular congruences (VRCs) and show that each VRC yields a VUT. Most examples of varieties of ranked tree languages are so-called principal varieties or unions of them (cf. [24, 25, 26]). A principal variety corresponds to a variety of congruences that consists of principal filters of the congruence lattices of term algebras in which the generating congruences are of finite index. Because of the unlimited branching in unranked trees, the corresponding VUTs cannot be defined this way. Instead, we introduce the notion of consistent systems of congruences.

These systems yield varieties of regular congruences and varieties of unranked tree languages that we call quasi-principal.

In Section 9 we establish a bijective correspondence between the varieties of regular algebras and the varieties of unranked tree languages. Section 10 contains several examples of VUTs that may be regarded as the natural unranked counterparts of some (general) varieties of ranked tree languages. Thus we have the VUTs of finite/co-finite, definite, reverse definite, generalized definite, aperiodic, locally testable and piecewise testable unranked tree languages. In most cases these VUTs are unions of quasi-principal VUTs defined by consistent systems of congruences. For example, for each  $k \geq 2$ , we define the VUT of  $k$ -testable unranked tree languages by means of a consistent system of congruences that is naturally given by the definition of  $k$ -testability, and the VUT of all locally testable unranked tree languages is the union of these VUTs.

In the final section we briefly review the results of the paper and note some further topics to be considered.

Several proofs that have been omitted or just outlined in the main text can be found in the Appendix at the end of the paper.

## 2 General preliminaries and unranked trees

We may write  $A := B$  to emphasize that  $A$  is defined to be  $B$ . Similarly,  $A :\Leftrightarrow B$  means that  $A$  is defined by the condition expressed by  $B$ . For any integer  $n \geq 0$ , let  $[n] := \{1, \dots, n\}$ . For a relation  $\rho \subseteq A \times B$ , the fact that  $(a, b) \in \rho$  is also expressed by  $a \rho b$  or  $a \equiv_\rho b$ . For any  $a \in A$ , let  $a\rho := \{b \mid a\rho b\}$ , and for  $A' \subseteq A$ , let  $A'\rho := \{b \in B \mid (\exists a \in A') a\rho b\}$ . The *converse* of  $\rho$  is the relation  $\rho^{-1} := \{(b, a) \mid a\rho b\} (\subseteq B \times A)$ . The *composition* of two relations  $\rho \subseteq A \times B$  and  $\rho' \subseteq B \times C$  is the relation  $\rho \circ \rho' := \{(a, c) \mid a \in A, c \in C, (\exists b \in B) a\rho b \text{ and } b\rho' c\}$ . The set of equivalence relations on a set  $A$  is denoted by  $\text{Eq}(A)$ , and for any  $\theta \in \text{Eq}(A)$ , let  $A/\theta := \{a\theta \mid a \in A\}$  be the corresponding quotient set. Let  $\Delta_A := \{(a, a) \mid a \in A\}$  be the *diagonal relation* and  $\nabla_A := A \times A$  be the *universal relation*.

For a mapping  $\varphi : A \rightarrow B$ , the image  $\varphi(a)$  of an element  $a \in A$  is also denoted by  $a\varphi$ . Accordingly, if  $H \subseteq A$  and  $K \subseteq B$ , we may also write  $H\varphi$  and  $K\varphi^{-1}$  for  $\varphi(H)$  and  $\varphi^{-1}(K)$ , respectively. Especially homomorphisms will be treated this way as right operators and the composition of  $\varphi : A \rightarrow B$  and  $\psi : B \rightarrow C$  is written as  $\varphi\psi$ . The *identity map*  $A \rightarrow A, a \mapsto a$ , is denoted by  $1_A$ . For any sets  $A_1, \dots, A_n$  ( $n \geq 1$ ) and any  $i \in [n]$ , we let  $\pi_i$  denote the  $i^{\text{th}}$  *projection*  $A_1 \times \dots \times A_n \rightarrow A_i, (a_1, \dots, a_n) \mapsto a_i$ .

For any alphabet  $X$ , we denote by  $X^*$  the set of all words over  $X$ , the empty word by  $\varepsilon$ , and by  $\text{Rec}(X^*)$  the set of all regular languages over  $X$ .

We need also some notions from lattice theory (cf. [7] or [10], for example). Let  $\leq$  be an order on a set  $L$ . A nonempty subset  $D$  of  $L$  is said to be *directed* if for all elements  $d_1, d_2 \in D$ , there is an element  $d \in D$  such that  $d_1, d_2 \leq d$ , and it is a *chain* in  $L$  if any two of its elements are comparable. Of course, any chain is directed. Now, let  $(L, \leq)$  be a lattice. A nonempty subset  $F \subseteq L$  is a

*filter* if

- (1)  $a, b \in F$  implies  $a \wedge b \in F$ , and
- (2)  $a \in F$ ,  $b \in L$  and  $a \leq b$  imply  $b \in F$ .

The *filter generated* by a nonempty subset  $H$  of  $L$ , i.e., the intersection of all filters containing  $H$ , is denoted by  $[H]$ . It is easy to see that

$$[H] = \{b \in L \mid a_1 \wedge \dots \wedge a_n \leq b \text{ for some } n \geq 1 \text{ and } a_1, \dots, a_n \in H\}.$$

As a special case, we get the *principal filter*  $[a] := \{b \in L \mid a \leq b\}$  generated by a singleton subset  $\{a\} \subseteq L$ .

The *unranked trees* to be considered here are finite and node-labeled, and their branches have a specified left-to-right order. Both from a theoretical point of view and for some applications, it will be natural to use two alphabets, an *operator alphabet* and a *leaf alphabet*, for labelling our trees. A symbol from the operator alphabet may appear as the label of any node of a tree, while the symbols of the leaf alphabet appear as labels of leaves only. In what follows,  $\Sigma$ ,  $\Omega$ ,  $\Gamma$  and  $\Psi$  denote operator alphabets, and  $X$ ,  $Y$  and  $Z$  leaf alphabets. We assume that all alphabets are finite and that operator alphabets are also nonempty.

**Definition 2.1.** The set  $T_\Sigma(X)$  of *unranked  $\Sigma X$ -trees* is the smallest set  $T$  of strings such that (1)  $X \cup \Sigma \subseteq T$ , and (2)  $f(t_1, \dots, t_m) \in T$  whenever  $f \in \Sigma$ ,  $m > 0$  and  $t_1, \dots, t_m \in T$ . Subsets of  $T_\Sigma(X)$  are called *unranked  $\Sigma X$ -tree languages*. Often we speak simply about  $\Sigma X$ -trees and  $\Sigma X$ -tree languages, or just about (unranked) trees and (unranked) tree languages without specifying the alphabets.  $\square$

Any  $u \in X \cup \Sigma$  represents a one-node tree in which the only node is labeled with  $u$ , and  $f(t_1, \dots, t_m)$  is interpreted as a tree formed by adjoining the  $m$  trees represented by  $t_1, \dots, t_m$  to a new  $f$ -labeled root in this left-to-right order.

As the definition of the set  $T_\Sigma(X)$  is inductive, notions relating to  $\Sigma X$ -trees can be defined recursively and statements about them can be proved by tree induction. For example, the *height*  $\text{hg}(t)$  and the *root (symbol)*  $\text{root}(t)$  of a  $\Sigma X$ -tree  $t$  are defined thus:

- (1)  $\text{hg}(u) = 0$  and  $\text{root}(u) = u$  for  $u \in \Sigma \cup X$ ;
- (2)  $\text{hg}(t) = \max\{\text{hg}(t_1), \dots, \text{hg}(t_m)\} + 1$  and  $\text{root}(t) = f$  for  $t = f(t_1, \dots, t_m)$ .

**Definition 2.2.** Let  $\xi$  be a special symbol not in  $\Sigma$  or  $X$ . A  $\Sigma X$ -context is a  $\Sigma(X \cup \{\xi\})$ -tree in which  $\xi$  appears exactly once. Let  $C_\Sigma(X)$  denote the set of all  $\Sigma X$ -contexts.

If  $p, q \in C_\Sigma(X)$ , then  $p(q)$  is the  $\Sigma X$ -context obtained from  $p$  by replacing the  $\xi$  in it with  $q$ . Similarly, if  $t \in T_\Sigma(X)$  and  $p \in C_\Sigma(X)$ , then  $p(t)$  is the  $\Sigma X$ -tree obtained when the  $\xi$  in  $p$  is replaced with  $t$ . The *height*  $\text{hg}(p)$  and the *root*  $\text{root}(p)$  of a  $\Sigma X$ -context  $p$  are defined treating  $p$  as a  $\Sigma(X \cup \{\xi\})$ -tree.  $\square$

Let us illustrate the above definitions by a few examples.

**Example 2.3.** Let  $f, g \in \Sigma$  and  $x, y \in X$ . The  $\Sigma X$ -tree  $t := f(g(y), x, f)$  and the  $\Sigma X$ -context  $p := f(\xi, f(g))$  are depicted in Figure 1. Now  $\text{hg}(t) = \text{hg}(p) = 2$ ,  $\text{root}(t) = \text{root}(p) = f$ ,  $p(t) = f(f(g(y), x, f), f(g))$  and  $p(g(\xi)) = f(g(\xi), f(g))$ . On the other hand,  $g(\xi)(p) = g(f(\xi, f(g)))$ .  $\square$

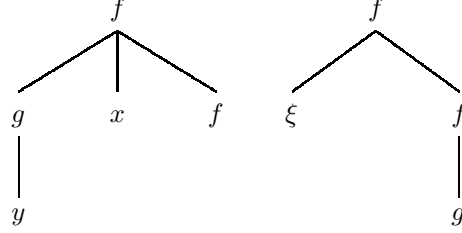


Figure 1:  $\Sigma X$ -tree  $f(g(y), x, f)$  and  $\Sigma X$ -context  $f(\xi, f(g))$

By the following examples we demonstrate that the use of two alphabets is quite natural also in typical applications of unranked trees.

**Example 2.4.** Figure 2 shows the tree representation of a small XML document. Here `invoices`, `invoice` and `line` belong to the operator alphabet while `text` is used as a generic name for a leaf symbol.  $\square$

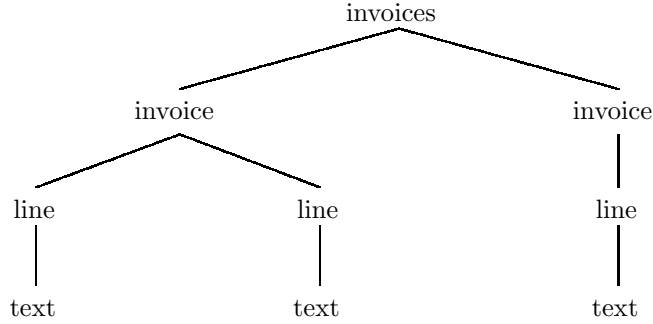


Figure 2: Unranked tree representing the structure of an XML document.

**Example 2.5.** In Example 3.2 of [17] (also Example 1 of [9]) the unranked trees are Boolean expressions without complements and variables, in which disjunctions and conjunctions may appear with any nonnegative arities. The alphabet consists of the symbols  $\vee, \wedge, 0$  and  $1$ . In our formalism it is natural to partition this set into an operator alphabet  $\Sigma = \{\vee, \wedge\}$  and a leaf alphabet  $X = \{0, 1\}$ ; the symbols  $0$  and  $1$  never label inner nodes.  $\square$

**Example 2.6.** Also the parse trees of sentences in a natural language are often best viewed as unranked. For example, in the tree shown in Figure 3, the label NP has multiple arities (2 and 3). The symbols S, VP, NP etc. that stand for the various grammatical categories form the operator alphabet while the words “This”, “singer”, “has” etc. belong to the leaf alphabet.  $\square$

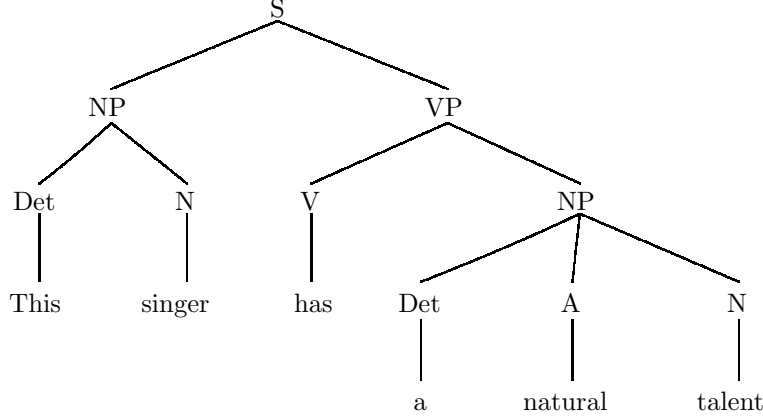


Figure 3: A parse tree of a sentence

### 3 Unranked algebras

For an algebraic theory of unranked tree languages we have to adapt the basic notions and facts of universal algebra to algebras over unranked sets of operation symbols. Since we will consider varieties containing tree languages over all alphabets, these notions are formulated in a way that corresponds to the generalized variety theory of [25]. The prefix *g* appearing in some names stands for “generalized”.

In the following, the set  $A$  of elements of an algebra will be regarded also as an alphabet and the set of all finite sequences of elements of  $A$  is denoted by  $A^*$ , and an  $m$ -tuple  $(a_1, \dots, a_m) \in A^m$  ( $m \geq 0$ ) may be written as the word  $a_1 \dots a_m$  and subsets of  $A^*$  may be viewed as languages.

**Definition 3.1.** An *unranked  $\Sigma$ -algebra*  $\mathcal{A}$  consists of a nonempty set  $A$  (of *elements* of  $\mathcal{A}$ ) and an operation  $f_{\mathcal{A}} : A^* \rightarrow A$  for each  $f \in \Sigma$ . We write simply  $\mathcal{A} = (A, \Sigma)$ . The algebra  $\mathcal{A}$  is *finite* if  $A$  is a finite set, and it is *trivial* if  $A$  is a one-element set. We may also speak just about  $\Sigma$ -algebras or *algebras* when there is no danger of confusion.  $\square$

These are essentially the “pseudoalgebras” used by Thatcher [29] who attributes the concept to J.R. Büchi and J.B. Wright (1960). In what follows,  $\mathcal{A} = (A, \Sigma)$ ,  $\mathcal{B} = (B, \Sigma)$ ,  $\mathcal{B} = (B, \Omega)$ ,  $\mathcal{C} = (C, \Gamma)$ , etc. are always unranked algebras with the operator alphabets shown.

**Definition 3.2.** If  $\Omega \subseteq \Sigma$ , an  $\Omega$ -algebra  $\mathcal{B} = (B, \Omega)$  is an  $\Omega$ -subalgebra of a  $\Sigma$ -algebra  $\mathcal{A} = (A, \Sigma)$  if  $B \subseteq A$  and  $f_{\mathcal{B}}(w) = f_{\mathcal{A}}(w)$  for all  $f \in \Omega$  and  $w \in B^*$ . Then we also call  $\mathcal{B}$  a  $g$ -subalgebra of  $\mathcal{A}$  without specifying  $\Omega$ . If  $\Omega = \Sigma$ , we call  $\mathcal{B}$  simply a subalgebra.  $\square$

If  $\mathcal{B} = (B, \Omega)$  is an  $\Omega$ -subalgebra of  $\mathcal{A} = (A, \Sigma)$ , then  $B$  is an  $\Omega$ -closed subset of  $\mathcal{A}$ , i.e.,  $f_{\mathcal{A}}(w) \in B$  for all  $f \in \Omega$  and  $w \in B^*$ . Any  $\Omega$ -closed subset  $B$  must be nonempty since  $\Omega \neq \emptyset$  and  $f_{\mathcal{A}}(\varepsilon) \in B$  for every  $f \in \Omega$ . This means that there is a perfect correspondence between  $\Omega$ -subalgebras and  $\Omega$ -closed subsets, and hence we may identify the two.

The intersection of any set of  $\Omega$ -subalgebras is an  $\Omega$ -subalgebra, and the  $\Omega$ -subalgebra  $\langle H \rangle_{\Omega}$  generated by a subset  $H \subseteq A$  can be defined in the usual way as the intersection of all  $\Omega$ -subalgebras that contain  $H$  as a subset. We write  $\langle H \rangle_{\Sigma}$  simply as  $\langle H \rangle$  and call it the subalgebra generated by  $H$ .

**Definition 3.3.** A pair of mappings  $\iota : \Sigma \rightarrow \Omega$ ,  $\varphi : A \rightarrow B$  forms a  $g$ -morphism from  $\mathcal{A} = (A, \Sigma)$  to  $\mathcal{B} = (B, \Omega)$ , written as  $(\iota, \varphi) : \mathcal{A} \rightarrow \mathcal{B}$ , if

$$f_{\mathcal{A}}(a_1, \dots, a_m)\varphi = \iota(f)_{\mathcal{B}}(a_1\varphi, \dots, a_m\varphi)$$

for all  $f \in \Sigma$ ,  $m \geq 0$  and  $a_1, \dots, a_m \in A$ . A  $g$ -morphism is a  $g$ -epimorphism, a  $g$ -monomorphism or a  $g$ -isomorphism if both maps are, respectively, surjective, injective or bijective. Two algebras  $\mathcal{A}$  and  $\mathcal{B}$  are said to be  $g$ -isomorphic,  $\mathcal{A} \cong_g \mathcal{B}$  in symbols, if there is a  $g$ -isomorphism  $(\iota, \varphi) : \mathcal{A} \rightarrow \mathcal{B}$ , and  $\mathcal{B}$  is called a  $g$ -image of  $\mathcal{A}$ , if there is a  $g$ -epimorphism from  $\mathcal{A}$  onto  $\mathcal{B}$ . Furthermore, an algebra  $\mathcal{A}$  is said to be  $g$ -covered by an algebra  $\mathcal{B}$ ,  $\mathcal{A} \preceq_g \mathcal{B}$  in symbols, if  $\mathcal{A}$  is a  $g$ -image of a  $g$ -subalgebra of  $\mathcal{B}$ .  $\square$

If  $(\iota, \varphi) : \mathcal{A} \rightarrow \mathcal{B}$  is a  $g$ -morphism as above, let  $\varphi_* : A^* \rightarrow B^*$  be the extension of  $\varphi$  to a monoid morphism. Then the fact that  $(\iota, \varphi)$  is a  $g$ -morphism can be expressed by saying that  $f_{\mathcal{A}}(w)\varphi = \iota(f)_{\mathcal{B}}(w\varphi_*)$  for all  $f \in \Sigma$  and  $w \in A^*$ .

A morphism  $\varphi : \mathcal{A} \rightarrow \mathcal{B}$  between two  $\Sigma$ -algebras  $\mathcal{A} = (A, \Sigma)$  and  $\mathcal{B} = (B, \Sigma)$  is a mapping  $\varphi : A \rightarrow B$  such that  $f_{\mathcal{A}}(w)\varphi = f_{\mathcal{B}}(w\varphi_*)$  for all  $f \in \Sigma$  and  $w \in A^*$ . It may be viewed as a special  $g$ -morphism  $(\iota, \varphi) : \mathcal{A} \rightarrow \mathcal{B}$  in which  $\iota$  is the identity map  $1_{\Sigma}$ . If  $\varphi$  is surjective, injective or bijective, then it is called an *epimorphism*, a *monomorphism* or an *isomorphism*, respectively. The algebras  $\mathcal{A}$  and  $\mathcal{B}$  are *isomorphic*,  $\mathcal{A} \cong \mathcal{B}$  in symbols, if there is an isomorphism  $\varphi : \mathcal{A} \rightarrow \mathcal{B}$ . Similarly,  $\mathcal{B}$  *covers*  $\mathcal{A}$ , and we express this by writing  $\mathcal{A} \preceq \mathcal{B}$ , if  $\mathcal{A}$  is an epimorphic image of some subalgebra of  $\mathcal{B}$ .

The  $g$ -morphisms of unranked algebras have all the basic properties of morphisms of ordinary algebras. Some of them are listed in the following lemma.

**Lemma 3.4.** Let  $\mathcal{A} = (A, \Sigma)$ ,  $\mathcal{B} = (B, \Omega)$  and  $\mathcal{C} = (C, \Gamma)$  be unranked algebras, and  $(\iota, \varphi) : \mathcal{A} \rightarrow \mathcal{B}$  and  $(\varkappa, \psi) : \mathcal{B} \rightarrow \mathcal{C}$  be  $g$ -morphisms.

- (a) The product  $(\iota\varkappa, \varphi\psi) : \mathcal{A} \rightarrow \mathcal{C}$  is also a  $g$ -morphism. Moreover, if  $(\iota, \varphi)$  and  $(\varkappa, \psi)$  are  $g$ -epi-,  $g$ -mono- or  $g$ -isomorphisms, then so is  $(\iota\varkappa, \varphi\psi)$ .



- (b) If  $R$  is a  $g$ -subalgebra of  $\mathcal{B}$ , then  $R\varphi^{-1}$  is a  $g$ -subalgebra of  $\mathcal{A}$ . In particular, if  $R$  is a  $\Psi$ -subalgebra of  $\mathcal{B}$  for some  $\Psi \subseteq \Omega$ , then  $R\varphi^{-1}$  is a  $\iota^{-1}(\Psi)$ -subalgebra of  $\mathcal{A}$ .
- (c) If  $S$  is a  $g$ -subalgebra of  $\mathcal{A}$ , then  $S\varphi$  is a  $g$ -subalgebra of  $\mathcal{B}$ . In particular, if  $S$  is a  $\Psi$ -subalgebra of  $\mathcal{A}$  for some  $\Psi \subseteq \Sigma$ , then  $S\varphi$  is a  $\iota(\Psi)$ -subalgebra of  $\mathcal{B}$ .

*Proof.* All of the assertions have straightforward proofs. As an example we consider statement (b). Let  $R$  be a  $\Psi$ -subalgebra of  $\mathcal{B}$  for some  $\Psi \subseteq \Omega$ . To show that  $R\varphi^{-1}$  is a  $\iota^{-1}(\Psi)$ -closed subset of  $\mathcal{A}$ , consider any  $f \in \iota^{-1}(\Psi)$ ,  $m \geq 0$  and  $a_1, \dots, a_m \in R\varphi^{-1}$ . Since  $R$  is  $\Psi$ -closed,  $\iota(f) \in \Psi$  and  $a_1\varphi, \dots, a_m\varphi \in R$ , we get

$$f_{\mathcal{A}}(a_1, \dots, a_m)\varphi = \iota(f)_{\mathcal{B}}(a_1\varphi, \dots, a_m\varphi) \in R,$$

and hence  $f_{\mathcal{A}}(a_1, \dots, a_m) \in R\varphi^{-1}$ .  $\square$

Next we consider congruences and quotients of unranked algebras.

**Definition 3.5.** A  $g$ -congruence on a  $\Sigma$ -algebra  $\mathcal{A} = (A, \Sigma)$  is a pair  $(\sigma, \theta)$ , where  $\sigma \in \text{Eq}(\Sigma)$  and  $\theta \in \text{Eq}(A)$ , such that for any  $f, g \in \Sigma$ ,  $m \geq 0$  and  $a_1, \dots, a_m, b_1, \dots, b_m \in A$ ,

$$f \sigma g, a_1 \theta b_1, \dots, a_m \theta b_m \Rightarrow f_{\mathcal{A}}(a_1, \dots, a_m) \theta g_{\mathcal{A}}(b_1, \dots, b_m).$$

Let  $\text{GCon}(\mathcal{A})$  denote the set of all  $g$ -congruences on  $\mathcal{A}$ .  $\square$

Any algebra  $\mathcal{A} = (A, \Sigma)$  has as  $g$ -congruences at least  $(\Delta_{\Sigma}, \Delta_A)$  and  $(\sigma, \nabla_A)$ , where  $\sigma$  is any equivalence on  $\Sigma$ . It is easy to see that with respect to the order defined by

$$(\sigma, \omega) \leq (\sigma', \omega') :\Leftrightarrow \sigma \subseteq \sigma' \text{ and } \omega \subseteq \omega' \quad ((\sigma, \omega), (\sigma', \omega') \in \text{GCon}(\mathcal{A})),$$

$\text{GCon}(\mathcal{A})$  forms a complete lattice in which joins and meets are formed componentwise in  $\text{Eq}(\Sigma)$  and  $\text{Eq}(A)$ , respectively.

The ordinary *congruences* of  $\mathcal{A} = (A, \Sigma)$  are the equivalences  $\theta \in \text{Eq}(A)$  such that  $(\Delta_{\Sigma}, \theta) \in \text{GCon}(\mathcal{A})$ . Their set is denoted by  $\text{Con}(\mathcal{A})$ . Note also that  $\theta \in \text{Con}(\mathcal{A})$  whenever  $(\omega, \theta) \in \text{GCon}(\mathcal{A})$  for some  $\omega \in \text{Eq}(\Sigma)$ .

**Definition 3.6.** For any  $g$ -congruence  $(\sigma, \theta) \in \text{GCon}(\mathcal{A})$  of an algebra  $\mathcal{A} = (A, \Sigma)$ , the  $g$ -quotient algebra  $\mathcal{A}/(\sigma, \theta) = (A/\theta, \Sigma/\sigma)$  is defined by setting

$$(f\sigma)_{\mathcal{A}/(\sigma, \theta)}(a_1\theta, \dots, a_m\theta) = f_{\mathcal{A}}(a_1, \dots, a_m)\theta$$

for all  $f \in \Sigma$ ,  $m \geq 0$ , and  $a_1, \dots, a_m \in A$ .  $\square$

The operations of  $\mathcal{A}/(\sigma, \theta)$  are clearly well-defined when  $(\sigma, \theta) \in \text{GCon}(\mathcal{A})$ . Note also that  $(f\sigma)_{\mathcal{A}/(\sigma, \theta)}(\varepsilon) = f_{\mathcal{A}}(\varepsilon)\theta$  for every  $f \in \Sigma$ .

The following lemma can be proved by appropriately modifying the usual proofs of the corresponding classical statements (cf. [7], for example).

**Lemma 3.7.** Let  $\mathcal{A} = (A, \Sigma)$  and  $\mathcal{B} = (B, \Omega)$  be any algebras.

- (a) For any  $g$ -congruence  $(\sigma, \theta)$  of  $\mathcal{A}$ , the maps  $\theta_{\natural} : A \rightarrow A/\theta, a \mapsto a\theta$ , and  $\sigma_{\natural} : \Sigma \rightarrow \Sigma/\sigma, f \mapsto f\sigma$ , define a  $g$ -epimorphism  $(\sigma_{\natural}, \theta_{\natural}) : \mathcal{A} \rightarrow \mathcal{A}/(\sigma, \theta)$ .
- (b) The kernel  $\ker(\iota, \varphi) := (\ker \iota, \ker \varphi)$  of a  $g$ -morphism  $(\iota, \varphi) : \mathcal{A} \rightarrow \mathcal{B}$  is a  $g$ -congruence of  $\mathcal{A}$ , and  $\mathcal{A}/\ker(\iota, \varphi) \cong_g \mathcal{B}$  if  $(\iota, \varphi)$  is a  $g$ -epimorphism.  $\square$

The quotient algebra  $\mathcal{A}/\theta = (A/\theta, \Sigma)$  of a unranked algebra  $\mathcal{A} = (A, \Sigma)$  with respect to a congruence  $\theta \in \text{Con}(\mathcal{A})$  is defined by setting  $f_{\mathcal{A}/\theta}(a_1\theta, \dots, a_m\theta) = f_{\mathcal{A}}(a_1, \dots, a_m)\theta$  for all  $f \in \Sigma, m \geq 0$  and  $a_1, \dots, a_m \in A$ . It may be regarded as a special  $g$ -quotient of  $\mathcal{A}$ ; if we identify in the natural way the operator alphabets  $\Sigma/\Delta_{\Sigma}$  and  $\Sigma$ , then  $\mathcal{A}/\theta$  is isomorphic to  $\mathcal{A}/(\Delta_{\Sigma}, \theta)$ .

**Definition 3.8.** For any mapping  $\varkappa : \Gamma \rightarrow \Sigma \times \Omega$ , the  $\varkappa$ -product of  $\mathcal{A} = (A, \Sigma)$  and  $\mathcal{B} = (B, \Omega)$  is the  $\Gamma$ -algebra  $\varkappa(\mathcal{A}, \mathcal{B}) = (A \times B, \Gamma)$  defined so that, for any  $f \in \Gamma, m \geq 0$  and  $(a_1, b_1), \dots, (a_m, b_m) \in A \times B$ ,

$$f_{\varkappa(\mathcal{A}, \mathcal{B})}((a_1, b_1), \dots, (a_m, b_m)) = (g_{\mathcal{A}}(a_1, \dots, a_m), h_{\mathcal{B}}(b_1, \dots, b_m)),$$

where  $(g, h) = \varkappa(f)$ . The products  $\varkappa(\mathcal{A}_1, \dots, \mathcal{A}_n)$  of any number  $n \geq 0$  of unranked algebras are defined similarly. Without specifying the map  $\varkappa$ , we call all such products jointly  $g$ -products. For  $n = 0$ , the  $g$ -product  $\varkappa(\mathcal{A}_1, \dots, \mathcal{A}_n)$  is taken to be the appropriate trivial algebra (unique up to isomorphism).  $\square$

The usual direct product  $\mathcal{A}_1 \times \dots \times \mathcal{A}_n$  of algebras  $\mathcal{A}_1 = (A_1, \Sigma), \dots, \mathcal{A}_n = (A_n, \Sigma)$  of the same type may be reconstructed as the  $g$ -product  $\varkappa(\mathcal{A}_1, \dots, \mathcal{A}_n)$  for  $\varkappa : \Sigma \rightarrow \Sigma \times \dots \times \Sigma, f \mapsto (f, \dots, f)$ . The  $g$ -products of just one factor are of special interest.

**Definition 3.9.** For any mapping  $\iota : \Sigma \rightarrow \Omega$  and any unranked  $\Omega$ -algebra  $\mathcal{B} = (B, \Omega)$ , the  $\Sigma$ -algebra  $\iota(\mathcal{B}) = (B, \Sigma)$  such that  $f_{\iota(\mathcal{B})} = \iota(f)_{\mathcal{B}}$  for every  $f \in \Sigma$ , is called the  $\iota$ -derived algebra of  $\mathcal{B}$  or, without specifying  $\iota$ , a  $g$ -derived algebra of  $\mathcal{B}$ .  $\square$

This notion is a natural analog of a special kind of the derived algebras considered in universal algebra (cf. [14, 20], and also [27]), and it has similar properties. In particular, we have the following obvious fact.

**Lemma 3.10.** If  $(\iota, \varphi) : \mathcal{A} \rightarrow \mathcal{B}$  is a  $g$ -morphism from a  $\Sigma$ -algebra  $\mathcal{A} = (A, \Sigma)$  to an  $\Omega$ -algebra  $\mathcal{B} = (B, \Omega)$ , then  $\varphi : \mathcal{A} \rightarrow \iota(\mathcal{B})$  is a morphism of  $\Sigma$ -algebras.  $\square$

Let us now consider subdirect decompositions of unranked algebras. Here it suffices to define all the appropriate notions with just finite algebras in mind. In the following,  $\Sigma_1, \dots, \Sigma_n$  and  $\Gamma$  are operator alphabets, and for the given map  $\varkappa : \Gamma \rightarrow \Sigma_1 \times \dots \times \Sigma_n$  and each  $i \in [n]$ , we denote by  $\varkappa_i$  the composition  $\varkappa\pi_i$  of  $\varkappa$  and the  $i^{\text{th}}$  projection  $\pi_i : \Sigma_1 \times \dots \times \Sigma_n \rightarrow \Sigma_i$ , i.e.,  $\varkappa_i(f) = \varkappa(f)\pi_i$  for every  $f \in \Gamma$ . Note that  $\pi_i$  also denotes the projection  $A_1 \times \dots \times A_n \rightarrow A_i$ .

**Definition 3.11.** A *generalized subdirect product*, a *gsd-product* for short, of some unranked algebras  $\mathcal{A}_1 = (A_1, \Sigma_1), \dots, \mathcal{A}_n = (A_n, \Sigma_n)$  ( $n \geq 0$ ) is a g-subalgebra  $\mathcal{B} = (B, \Omega)$  of a g-product  $\kappa(\mathcal{A}_1, \dots, \mathcal{A}_n) = (A_1 \times \dots \times A_n, \Gamma)$  such that  $B\pi_i = A_i$  and  $\Omega\kappa_i = \Sigma_i$  for every  $i \in [n]$ .

A *gsd-representation* of  $\mathcal{A} = (A, \Sigma)$ , with factors  $\mathcal{A}_1 = (A_1, \Sigma_1), \dots, \mathcal{A}_n = (A_n, \Sigma_n)$  ( $n \geq 0$ ), is a g-monomorphism  $(\iota, \varphi) : \mathcal{A} \rightarrow \kappa(\mathcal{A}_1, \dots, \mathcal{A}_n)$ , where also  $\kappa : \Gamma \rightarrow \Sigma_1 \times \dots \times \Sigma_n$  is injective, such that  $A\varphi\pi_i = A_i$  and  $\Sigma\iota\kappa_i = \Sigma_i$  for every  $i \in [n]$ . Such a gsd-representation is *proper* if for no  $i \in [n]$ , both  $\varphi\pi_i : A \rightarrow A_i$  and  $\iota\kappa_i : \Sigma \rightarrow \Sigma_i$  are injective. A finite unranked algebra is *gsd-irreducible* if it has no proper gsd-representation.  $\square$

That the above gsd-representation is proper means that none of the g-epimorphisms  $(\iota\kappa_i, \varphi\pi_i) : \mathcal{A} \rightarrow \mathcal{A}_i$  is a g-isomorphism (cf. Lemma 3.13 below).

**Remark 3.12.** If we compose the maps  $\iota : \Sigma \rightarrow \Gamma$  and  $\kappa : \Gamma \rightarrow \Sigma_1 \times \dots \times \Sigma_n$  of the above definition, the gsd-representation  $(\iota, \varphi) : \mathcal{A} \rightarrow \kappa(\mathcal{A}_1, \dots, \mathcal{A}_n)$  yields a monomorphism  $\varphi : \mathcal{A} \rightarrow (\iota\kappa)(\mathcal{A}_1, \dots, \mathcal{A}_n)$  of  $\Sigma$ -algebras.  $\square$

The next two lemmas show the links between the gsd-representations (with finitely many factors) and the g-congruences of an unranked algebra. They can be proved in the same way as their classical counterparts (cf. [3], for example).

**Lemma 3.13.** Let  $(\iota, \varphi) : \mathcal{A} \rightarrow \kappa(\mathcal{A}_1, \dots, \mathcal{A}_n)$  be a gsd-representation of an unranked algebra  $\mathcal{A} = (A, \Sigma)$  with factors  $\mathcal{A}_1 = (A_1, \Sigma_1), \dots, \mathcal{A}_n = (A_n, \Sigma_n)$  ( $n \geq 0$ ). For every  $i \in [n]$ ,  $(\iota\kappa_i, \varphi\pi_i) : \mathcal{A} \rightarrow \mathcal{A}_i$  is a g-epimorphism. Moreover, if we write  $(\sigma_i, \theta_i) := \ker(\iota\kappa_i, \varphi\pi_i)$  for each  $i \in [n]$ , then

- (a)  $(\sigma_i, \theta_i) \in \text{GCon}(\mathcal{A})$  and  $\mathcal{A}/(\sigma_i, \theta_i) \cong_g \mathcal{A}_i$ , and
- (b)  $(\sigma_1, \theta_1) \wedge \dots \wedge (\sigma_n, \theta_n) = (\Delta_\Sigma, \Delta_A)$ .

If the representation is proper, then  $(\sigma_i, \theta_i) > (\Delta_\Sigma, \Delta_A)$  for every  $i \in [n]$ .  $\square$

**Lemma 3.14.** If an unranked algebra  $\mathcal{A} = (A, \Sigma)$  has g-congruences  $(\sigma_1, \theta_1), \dots, (\sigma_n, \theta_n)$  such that  $(\sigma_1, \theta_1) \wedge \dots \wedge (\sigma_n, \theta_n) = (\Delta_\Sigma, \Delta_A)$ , then

$$(1_\Sigma, \varphi) : \mathcal{A} \rightarrow \kappa(\mathcal{A}/(\sigma_1, \theta_1), \dots, \mathcal{A}/(\sigma_n, \theta_n))$$

is a gsd-representation of  $\mathcal{A}$  for  $\kappa : \Sigma \rightarrow \Sigma/\sigma_1 \times \dots \times \Sigma/\sigma_n$ ,  $f \mapsto (f\sigma_1, \dots, f\sigma_n)$ , and  $\varphi : A \rightarrow A/\theta_1 \times \dots \times A/\theta_n$ ,  $a \mapsto (a\theta_1, \dots, a\theta_n)$ . If  $(\sigma_i, \theta_i) > (\Delta_\Sigma, \Delta_A)$  for every  $i \in [n]$ , then this gsd-representation is proper.  $\square$

The following proposition contains the counterparts of Birkhoff's two fundamental theorems about subdirect representations.

**Proposition 3.15.** Let  $\mathcal{A} = (A, \Sigma)$  be a finite unranked algebra.

- (a)  $\mathcal{A}$  is gsd-irreducible if and only if  $|A| = |\Sigma| = 1$  or it has a least nontrivial g-congruence, i.e.,  $\bigcap (\text{GCon}(\mathcal{A}) \setminus \{(\Delta_\Sigma, \Delta_A)\}) > (\Delta_\Sigma, \Delta_A)$ .

- (b)  $\mathcal{A}$  has a *gsd-representation with finitely many factors each of which is a gsd-irreducible  $g$ -image of  $\mathcal{A}$* .

*Proof.* Statement (a) follows from Lemmas 3.13 and 3.14. Statement (b) could be proved similarly as its classical counterpart (cf. [7], for example).  $\square$

Note that a trivial  $\Sigma$ -algebra  $\mathcal{A} = (A, \Sigma)$  is gsd-irreducible exactly in case  $|\Sigma| \leq 2$ . Indeed, if  $|\Sigma| > 2$ , then  $\mathcal{A}$  has a proper gsd-representation in which the factors are trivial algebras with smaller operator alphabets. If  $|\Sigma| = 2$ , then  $(\nabla_\Sigma, \Delta_A)$  is the least nontrivial  $g$ -congruence of  $\mathcal{A}$ .

**Definition 3.16.** For any  $\Sigma$  and  $X$ , we define the *unranked  $\Sigma X$ -term algebra*  $\mathcal{T}_\Sigma(X) = (T_\Sigma(X), \Sigma)$  in such a way that for any  $f \in \Sigma$ ,

- (1)  $f_{\mathcal{T}_\Sigma(X)}(\varepsilon) = f$ , and
- (2)  $f_{\mathcal{T}_\Sigma(X)}(t_1, \dots, t_m) = f(t_1, \dots, t_m)$  for any  $m > 0$  and  $t_1, \dots, t_m \in T_\Sigma(X)$ .

Again, we may speak simply about the  $\Sigma X$ -term algebra or a *term algebra*.  $\square$

A  $g$ -morphism  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{T}_\Omega(Y)$  between unranked term algebras defines a mapping from  $T_\Sigma(X)$  to  $T_\Omega(Y)$  that replaces each label  $f \in \Sigma$  with  $\iota(f) \in \Omega$  and each leaf labeled with a symbol  $x \in X$  with the  $\Omega Y$ -tree  $x\varphi$ . Such mappings are the unranked analogs of the *inner alphabetic tree homomorphisms* considered in [15] (in the form they appear in [25]). For any given  $\iota : \Sigma \rightarrow \Omega$  and any leaf alphabet  $X$ , we define a special mapping  $\iota_X : T_\Sigma(X) \rightarrow T_\Omega(X)$  of this type as follows:

- (1)  $\iota_X(x) = x$  for  $x \in X$ ;
- (2)  $\iota_X(f) = \iota(f)$  for  $f \in \Sigma$ ;
- (3)  $\iota_X(t) = \iota(f)(\iota_X(t_1), \dots, \iota_X(t_m))$  for  $t = f(t_1, \dots, t_m)$ .

Then  $(\iota, \iota_X) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{T}_\Omega(X)$  is a  $g$ -morphism that transforms any  $\Sigma X$ -tree to an  $\Omega X$ -tree by replacing any label  $f \in \Sigma$  with  $\iota(f)$  but preserving all symbols from  $X$ .

The following proposition expresses an important property of our term algebras.

**Proposition 3.17.** *For any  $\Sigma$  and any  $X$ , the term algebra  $\mathcal{T}_\Sigma(X)$  is freely generated by  $X$  over the class of all unranked algebras, that is to say,*

- (a)  $\langle X \rangle = T_\Sigma(X)$ , and
- (b) *if  $\mathcal{A} = (A, \Omega)$  is any unranked algebra, then for any pair of mappings  $\iota : \Sigma \rightarrow \Omega$  and  $\alpha : X \rightarrow A$ , there is a unique  $g$ -morphism  $(\iota, \varphi_{\iota, \alpha}) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{A}$  such that  $\varphi_{\iota, \alpha}|_X = \alpha$ .*

*Proof.* (a) It is clear that  $X \subseteq \langle X \rangle \subseteq T_\Sigma(X)$ , and that to prove  $T_\Sigma(X) \subseteq \langle X \rangle$ , it suffices to show that  $T_\Sigma(X) \subseteq B$  for any  $\Sigma$ -closed subset  $B$  of  $T_\Sigma(X)$  for which  $X \subseteq B$ . This can be done simply by tree induction.

(b) For any  $\iota : \Sigma \rightarrow \Omega$  and  $\alpha : X \rightarrow A$ , a  $g$ -morphism  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{A}$  such that  $\varphi|_X = \alpha$  must satisfy the following conditions:

- (1)  $x\varphi = \alpha(x)$  for  $x \in X$ ;
- (2)  $f\varphi = f_{\mathcal{T}_\Sigma(X)}(\varepsilon)\varphi = \iota(f)\mathcal{A}(\varepsilon)$  for  $f \in \Sigma$ ;
- (3)  $t\varphi = \iota(f)\mathcal{A}(t_1\varphi, \dots, t_m\varphi)$  for  $t = f(t_1, \dots, t_m)$ .

It is easy to show by tree induction that these conditions assign a unique value  $t\varphi$  to each  $t \in T_\Sigma(X)$ , and that the thus defined  $(\iota, \varphi)$  is a  $g$ -morphism.  $\square$

Similarly as for ordinary algebras, the values of  $\varphi_{\iota, \alpha}$  can be obtained by evaluating term functions for the valuation  $\alpha : X \rightarrow A$ .

**Definition 3.18.** For any operator alphabet  $\Sigma$  and leaf alphabet  $X$ , the *term function*  $t^{\mathcal{A}} : A^X \rightarrow A$  of a  $\Sigma$ -algebra  $\mathcal{A} = (A, \Sigma)$  induced by a  $\Sigma X$ -tree  $t \in T_\Sigma(X)$  is defined as follows. For any  $\alpha : X \rightarrow A$ ,

- (1)  $x^{\mathcal{A}}(\alpha) = \alpha(x)$  for any  $x \in X$
- (2)  $f^{\mathcal{A}}(\alpha) = f_{\mathcal{A}}(\varepsilon)$  for any  $f \in \Sigma$ , and
- (3)  $t^{\mathcal{A}}(\alpha) = f_{\mathcal{A}}(t_1^{\mathcal{A}}(\alpha), \dots, t_m^{\mathcal{A}}(\alpha))$  for  $t = f(t_1, \dots, t_m)$ .  $\square$

It is easy to see that with  $\mathcal{A}$ ,  $\iota$  and  $\alpha$  as in Proposition 3.17,  $t\varphi_{\iota, \alpha} = \iota_X(t)^{\mathcal{A}}(\alpha)$  for every  $t \in T_\Sigma(X)$ . The following notion will also be useful.

**Definition 3.19.** A mapping  $p : A \rightarrow A$  is an *elementary translation* of an unranked  $\Sigma$ -algebra  $\mathcal{A} = (A, \Sigma)$  if there exist an  $f \in \Sigma$ , an  $m > 0$ , and an  $i \in [m]$  and elements  $a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_m \in A$  such that  $p(b) = f_{\mathcal{A}}(a_1 \dots a_{i-1} b a_{i+1} \dots a_m)$  for all  $b \in A$ . The set  $\text{Tr}(\mathcal{A})$  of *translations* of  $\mathcal{A}$  is defined as the smallest set of mappings  $A \rightarrow A$  that contains the identity map  $1_A : A \rightarrow A$  and all elementary translations of  $\mathcal{A}$ , and is closed under composition.  $\square$

The following facts can be shown exactly as for ordinary algebras (cf. [7], for example).

**Lemma 3.20.** *Any congruence of an unranked algebra  $\mathcal{A} = (A, \Sigma)$  is invariant with respect to all translations of  $\mathcal{A}$ , and an equivalence on  $A$  is a congruence of  $\mathcal{A}$  if it is invariant with respect to all elementary translations of  $\mathcal{A}$ .*  $\square$

Moreover, we have the following counterpart to Lemma 5.3 of [25].

**Lemma 3.21.** *Let  $(\iota, \varphi) : \mathcal{A} \rightarrow \mathcal{B}$  be a  $g$ -morphism from a  $\Sigma$ -algebra  $\mathcal{A} = (A, \Sigma)$  to an  $\Omega$ -algebra  $\mathcal{B} = (B, \Omega)$ . For every translation  $p \in \text{Tr}(\mathcal{A})$ , there is a translation  $p_{\iota, \varphi}$  of  $\mathcal{B}$  such that  $p(a)\varphi = p_{\iota, \varphi}(a\varphi)$  for every  $a \in A$ . If  $(\iota, \varphi)$  is a  $g$ -epimorphism, then every translation of  $\mathcal{B}$  equals  $p_{\iota, \varphi}$  for some  $p \in \text{Tr}(\mathcal{A})$ .  $\square$*

Finally, it is easy to see that the translations of a term algebra  $\mathcal{T}_\Sigma(X)$  can be defined by and correspond bijectively to the  $\Sigma X$ -contexts: for every  $p \in \text{Tr}(\mathcal{T}_\Sigma(X))$  there is a unique  $q \in C_\Sigma(X)$  such that  $p(t) = q(t)$  for every  $t \in T_\Sigma(X)$ , and conversely.

## 4 Regular unranked algebras

Let us now introduce the unranked algebras that will play the same role here as finite algebras in the ranked case. In [29] they were called “pseudoautomata”.

**Definition 4.1.** An unranked algebra  $\mathcal{A} = (A, \Sigma)$  is said to be *regular* if it is finite and  $f_{\mathcal{A}}^{-1}(a)$  is a regular language over  $A$  for all  $f \in \Sigma$  and  $a \in A$ .  $\square$

The condition that all the sets  $f_{\mathcal{A}}^{-1}(a) = \{w \in A^* \mid f_{\mathcal{A}}(w) = a\}$  be regular languages means that the functions  $f_{\mathcal{A}} : A^* \rightarrow A$  can be computed by finite automata.

**Example 4.2.** Let  $\Sigma = \{f\}$  and let  $\mathcal{A} = (\{0, 1\}, \Sigma)$  be the unranked  $\Sigma$ -algebra such that  $f_{\mathcal{A}}(w) = 1$  if  $w \in \{0^n 1^n \mid n \geq 0\}$ , and  $f_{\mathcal{A}}(w) = 0$  otherwise ( $w \in A^*$ ). In this case  $\mathcal{A}$  is not regular since the language  $f_{\mathcal{A}}^{-1}(1) = \{0^n 1^n \mid n \geq 0\}$  is not regular. On the other hand, a regular algebra  $\mathcal{A}$  is obtained if we define  $f_{\mathcal{A}}$  by setting

$$f_{\mathcal{A}}(w) = a_1 + \dots + a_n \pmod{2},$$

for  $w = a_1 \dots a_n$ , where  $a_1, \dots, a_n \in A$ .  $\square$

**Lemma 4.3.** All  $g$ -subalgebras and all  $g$ -images of a regular algebra are regular.

*Proof.* Let  $\mathcal{A} = (A, \Sigma)$  be a regular algebra and let  $\mathcal{B} = (B, \Omega)$  be any unranked algebra.

If  $\mathcal{B}$  is an  $\Omega$ -subalgebra of  $\mathcal{A}$ , then  $f_{\mathcal{B}}^{-1}(b) = f_{\mathcal{A}}^{-1}(b) \cap B^*$  is a regular language for all  $f \in \Omega$  and  $b \in B$ , and hence  $\mathcal{B}$  is regular.

Assume now that there is a  $g$ -epimorphism  $(\iota, \varphi) : \mathcal{A} \rightarrow \mathcal{B}$ . Consider any  $g \in \Omega$ ,  $b \in B$  and  $w \in B^*$ . If  $\varphi_* : A^* \rightarrow B^*$  is the extension of  $\varphi$  to a monoid morphism, then  $w = v\varphi_*$  for some  $v \in A^*$ . If  $f \in \Sigma$  is such that  $\iota(f) = g$ , then

$$\begin{aligned} w \in g_{\mathcal{B}}^{-1}(b) &\Leftrightarrow g_{\mathcal{B}}(w) = b \Leftrightarrow \iota(f)_{\mathcal{B}}(v\varphi_*) = b \Leftrightarrow f_{\mathcal{A}}(v)\varphi = b \Leftrightarrow v \in f_{\mathcal{A}}^{-1}(b\varphi^{-1}) \\ &\Rightarrow w \in f_{\mathcal{A}}^{-1}(b\varphi^{-1})\varphi_*, \end{aligned}$$

i.e.,  $g_{\mathcal{B}}^{-1}(b) \subseteq f_{\mathcal{A}}^{-1}(b\varphi^{-1})\varphi_*$ . For the converse inclusion, let  $w \in f_{\mathcal{A}}^{-1}(b\varphi^{-1})\varphi_*$ . Then  $w = v\varphi_*$  for some  $a \in b\varphi^{-1}$  and  $v \in f_{\mathcal{A}}^{-1}(a)$ . This means that

$$b = a\varphi = f_{\mathcal{A}}(v)\varphi = g_{\mathcal{B}}(v\varphi_*) = g_{\mathcal{B}}(w),$$

and hence  $w \in g_{\mathcal{B}}^{-1}(b)$ . Since  $f_{\mathcal{A}}^{-1}(b\varphi^{-1})$  is the union of finitely many regular sets  $f_{\mathcal{A}}^{-1}(a)$ , where  $a \in b\varphi^{-1}$ , this means that  $g_{\mathcal{B}}^{-1}(b)$  is regular.  $\square$

**Lemma 4.4.** Any  $g$ -product of regular algebras is regular. In particular, any  $g$ -derived algebra of a regular algebra is regular.

*Proof.* To simplify the notation, we consider a g-product  $\varkappa(\mathcal{A}, \mathcal{B}) = (A \times B, \Gamma)$  of just two regular algebras  $\mathcal{A} = (A, \Sigma)$  and  $\mathcal{B} = (B, \Omega)$ . Let  $f \in \Gamma$  and  $(a, b) \in A \times B$ , and assume that  $\varkappa(f) = (g, h)$ . If  $\varphi_1 : (A \times B)^* \rightarrow A^*$  and  $\varphi_2 : (A \times B)^* \rightarrow B^*$  are the extensions of the projections  $\pi_1 : A \times B \rightarrow A$  and  $\pi_2 : A \times B \rightarrow B$  to monoid morphisms, then  $f_{\varkappa(\mathcal{A}, \mathcal{B})}(w) = (g_{\mathcal{A}}(w\varphi_1), h_{\mathcal{B}}(w\varphi_2))$  for any  $w \in (A \times B)^*$ . From this it follows that  $f_{\varkappa(\mathcal{A}, \mathcal{B})}^{-1}(a, b) = g_{\mathcal{A}}^{-1}(a)\varphi_1^{-1} \cap h_{\mathcal{B}}^{-1}(b)\varphi_2^{-1}$ , which is a regular set.  $\square$

Let us say that a regular algebra  $\mathcal{A} = (A, \Sigma)$  is *effectively given* if, for all  $f \in \Sigma$  and  $a \in A$ , we are given a finite recognizer of  $f_{\mathcal{A}}^{-1}(a)$ . The following proposition expresses an important property of regular algebras.

**Proposition 4.5.** *For any effectively given regular algebra  $\mathcal{A} = (A, \Sigma)$ , the set  $\text{Tr}(\mathcal{A})$  of all translations of  $\mathcal{A}$  is effectively computable.*

*Proof.* For any  $f \in \Sigma$  and  $u, v \in A^*$ , let  $f_{u,v} : A \rightarrow A, a \mapsto f_{\mathcal{A}}(uav)$ , be the elementary translation defined by  $f$ ,  $u$  and  $v$ . We show that  $E_f := \{f_{u,v} \mid u, v \in A^*\}$  is effectively computable for each  $f$ .

For each  $a \in A$ , we can find a finite monoid  $M_a$ , a morphism  $\varphi_a : A^* \rightarrow M_a$  and a subset  $F_a \subseteq M_a$  such that  $f_{\mathcal{A}}^{-1}(a) = F_a\varphi_a^{-1}$ . Let  $\sim$  be the equivalence on  $A^*$  such that for any  $u, v \in A^*$ ,  $u \sim v$  if and only if  $u\varphi_a = v\varphi_a$  for every  $a \in A$ . Then  $f_{u,v} = f_{u',v'}$  for all words  $u, v, u', v' \in A^*$  such that  $u \sim u'$  and  $v \sim v'$ . Indeed, for all  $a, b \in A$ ,

$$\begin{aligned} f_{u,v}(a) = b &\Leftrightarrow f_{\mathcal{A}}(uav) = b \Leftrightarrow uav \in f_{\mathcal{A}}^{-1}(b) \Leftrightarrow (uav)\varphi_b \in F_b \\ &\Leftrightarrow u\varphi_b \cdot a\varphi_b \cdot v\varphi_b \in F_b \Leftrightarrow u'\varphi_b \cdot a\varphi_b \cdot v'\varphi_b \in F_b \Leftrightarrow \dots \\ &\Leftrightarrow f_{u',v'}(a) = b. \end{aligned}$$

Let  $R$  be a set of representatives of the partition  $A^*/\sim$ . Such an  $R$  is finite and can be effectively formed using the regular sets  $m\varphi_a^{-1}$  ( $m \in M_a, a \in A$ ). Then  $E_f$  is obtained as the set  $\{f_{u,v} \mid u, v \in R\}$ .  $\square$

We shall consider classes containing unranked  $\Sigma$ -algebras for any operator alphabet  $\Sigma$ . Note that also the operators  $S$ ,  $H$  and  $P_f$  are now applied to such classes. The class of all regular algebras is denoted by **Reg**.

**Definition 4.6.** For any class **K** of unranked algebras, let

- (1)  $S_g(\mathbf{K})$  be the class of algebras g-isomorphic to a g-subalgebra of a member **K**,
- (2)  $H_g(\mathbf{K})$  be the class of all g-images of members of **K**,
- (3)  $P_g(\mathbf{K})$  be the class of algebras isomorphic to g-products of members of **K**,
- (4)  $S(\mathbf{K})$  be the class of algebras isomorphic to a subalgebra of a member **K**,
- (5)  $H(\mathbf{K})$  be the class of all epimorphic images of members of **K**, and

- (6)  $P_f(\mathbf{K})$  be the class of algebras isomorphic to the direct product of a finite family of members of  $\mathbf{K}$ .

A class  $\mathbf{K}$  of regular unranked algebras is a *variety of regular algebras (VRA)* if  $S_g(\mathbf{K}), H_g(\mathbf{K}), P_g(\mathbf{K}) \subseteq \mathbf{K}$ . The class of all VRAs is denoted by **VRA**.  $\square$

Since g-derived algebras are special g-products, the following fact is an immediate consequence of the definition of VRAs.

**Lemma 4.7.** *Every VRA is closed under the forming of g-derived algebras.*  $\square$

From Lemmas 4.3 and 4.4 we get the following proposition.

**Proposition 4.8.** *The class **Reg** of all regular unranked algebras is a VRA, and hence it is the greatest VRA.*  $\square$

The intersection of all VRAs that contain a given class  $\mathbf{K}$  of regular algebras is obviously a VRA. It is called the *VRA generated by  $\mathbf{K}$*  and it is denoted by  $V_g(\mathbf{K})$ .

If  $P$  and  $Q$  are any algebra class operators, such as  $H_g, S_g$  or  $P_f$ , we denote, as usual, by  $PQ$  the operator such that for any class  $\mathbf{K}$  of algebras,  $PQ(\mathbf{K}) = P(Q(\mathbf{K}))$ . Moreover,  $P \leq Q$  means that  $P(\mathbf{K}) \subseteq Q(\mathbf{K})$  for every class  $\mathbf{K}$ .

The obvious fact that  $\mathbf{K} \subseteq S(\mathbf{K}) \subseteq S_g(\mathbf{K})$ ,  $\mathbf{K} \subseteq H(\mathbf{K}) \subseteq H_g(\mathbf{K})$ , and  $\mathbf{K} \subseteq P_f(\mathbf{K}) \subseteq P_g(\mathbf{K})$  for any class  $\mathbf{K}$  of finite unranked algebras, will be used without comment.

To obtain an analog of Tarski's HSP-representation of the generated variety-operator (cf. [7] or [3]) for  $V_g$ , we first prove some commutation and semi-commutation relations of our class operators.

**Lemma 4.9.** (a)  $S_g S_g = S_g S = S S_g = S_g$ .

(b)  $H_g H_g = H_g H = H H_g = H_g$ .

(c)  $P_g P_g = P_g P_f = P_f P_g = P_g$ .

(d)  $S_g H \leq S_g H_g \leq H S_g \leq H_g S_g$ .

(e)  $P_g S \leq P_g S_g \leq S P_g \leq S_g P_g$ .

(f)  $P_g H \leq P_g H_g \leq H P_g \leq H_g P_g$ .

*Proof.* Statements (a) and (b) hold because obviously  $S_g S_g = S_g$  and  $H_g H_g = H_g$ . To prove (c), it clearly suffices to show that  $P_g P_g \leq P_g$ , and in each of (d), (e) and (f), it suffices to prove the second inequality because the other two are obvious. By the way of an example, we verify the inequality  $S_g H_g \leq H S_g$ .

Let  $\mathbf{K}$  be any class of unranked algebras. To construct a typical member  $\mathcal{C} = (C, \Gamma)$  of  $S_g H_g(\mathbf{K})$ , let  $\mathcal{A} = (A, \Sigma)$  be in  $\mathbf{K}$ ,  $(\iota, \varphi) : \mathcal{A} \rightarrow \mathcal{A}'$  be a g-epimorphism,  $\mathcal{B} = (B, \Omega)$  be a g-subalgebra of  $\mathcal{A}'$ , and let  $(\varkappa, \psi) : \mathcal{B} \rightarrow \mathcal{C}$  be a g-isomorphism. Now  $\mathcal{B}\varphi^{-1} = (B\varphi^{-1}, \iota^{-1}(\Omega))$  is a g-subalgebra of  $\mathcal{A}$ . If we choose a subset  $\Sigma'$  of  $\iota^{-1}(\Omega)$  so that the restriction of  $\iota$  to  $\Sigma'$  is a bijection  $\iota' : \Sigma' \rightarrow \Omega$ , then  $\mathcal{D} = (B\varphi^{-1}, \Sigma')$  is a g-subalgebra of  $\mathcal{A}$ .



Next we define a  $\Gamma$ -algebra  $\mathcal{E} = (B\varphi^{-1}, \Gamma)$  so that for each  $g \in \Gamma$ ,  $g_{\mathcal{E}} = f_{\mathcal{D}}$  for the  $f \in \Sigma'$  such that  $g = \varkappa(\iota'(f))$ . Then  $(\iota'\varkappa, 1_{B\varphi^{-1}}) : \mathcal{D} \rightarrow \mathcal{E}$  is a g-isomorphism. Indeed, if  $f \in \Sigma'$  and  $w \in (B\varphi^{-1})^*$ , then

$$f_{\mathcal{D}}(w)1_{B\varphi^{-1}} = f_{\mathcal{D}}(w) = \varkappa(\iota'(f))_{\mathcal{E}}(w1_{B\varphi^{-1}}).$$

This means that  $\mathcal{E} \in S_g(\mathbf{K})$ . Next, we show that  $\varphi\psi : \mathcal{E} \rightarrow \mathcal{C}$  is an epimorphism. Clearly,  $B\varphi^{-1}\varphi\psi = C$ . Consider any  $g \in \Gamma$  and  $w \in (B\varphi^{-1})^*$ . Let  $f \in \Sigma'$  and  $h \in \Omega$  be such that  $\iota'(f) = h$  and  $\varkappa(h) = g$ . Then

$$g_{\mathcal{E}}(w)\varphi\psi = f_{\mathcal{D}}(w)\varphi\psi = f_{\mathcal{A}}(w)\varphi\psi = h_{\mathcal{A}'}(w\varphi)\psi = h_{\mathcal{B}}(w\varphi)\psi = g_{\mathcal{C}}(w\varphi\psi).$$

Thus  $\mathcal{C} \in HS_g(\mathbf{K})$ .  $\square$

The relations  $S_g S_g = S_g$ ,  $H_g H_g = H_g$ ,  $P_g P_g = P_g$ ,  $S_g H_g \leq H_g S_g$ ,  $P_g S_g \leq S_g P_g$ , and  $P_g H_g \leq H_g P_g$  yield, in the usual way (cf. [7] or [3], for example) the following representation of the  $V_g$ -operator.

**Proposition 4.10.**  $V_g = H_g S_g P_g$ .  $\square$

For a simpler representation of the  $V_g$ -operator, in which just the  $P$ -operator appears in the generalized form, we need also the following two relations.

**Lemma 4.11.** (a)  $H_g S \leq HS_g$ , and (b)  $S_g P_g \leq SP_g$ .

*Proof.* Throughout this proof,  $\mathbf{K}$  is again any given class of unranked algebras.

For proving (a), let  $\mathcal{A} = (A, \Sigma) \in \mathbf{K}$ ,  $\mathcal{B} = (B, \Sigma)$  be a subalgebra of  $\mathcal{A}$ , and let  $(\iota, \varphi) : \mathcal{B} \rightarrow \mathcal{C}$  be a g-epimorphism onto a  $\Gamma$ -algebra  $\mathcal{C} = (C, \Gamma)$ .

Let  $\Omega \subseteq \Sigma$  be such that the restriction  $\iota' : \Omega \rightarrow \Gamma$  of  $\iota$  to  $\Omega$  is a bijection. Then  $\mathcal{B}' = (B, \Omega)$  is a g-subalgebra of  $\mathcal{A}$ . Define a new algebra  $\mathcal{B}'' = (B, \Gamma)$  by setting for each  $g \in \Gamma$ ,  $g_{\mathcal{B}''} = h_{\mathcal{B}'}$  for the  $h \in \Omega$  such that  $\iota'(h) = g$ . It is easy to show that  $(\iota', 1_B) : \mathcal{B}' \rightarrow \mathcal{B}''$  is a g-isomorphism. Hence,  $\mathcal{B}'' \in S_g(\mathbf{K})$ . To prove  $\mathcal{C} \in HS_g(\mathbf{K})$ , it suffices to verify that  $\varphi : \mathcal{B}'' \rightarrow \mathcal{C}$  is an epimorphism of  $\Gamma$ -algebras. Consider any  $g \in \Gamma$ ,  $m \geq 0$ , and  $b_1, \dots, b_m \in B$ , and let  $h \in \Omega$  be the symbol such that  $\iota(h) = g$ . Then

$$g_{\mathcal{B}''}(b_1, \dots, b_m)\varphi = h_{\mathcal{B}'}(b_1, \dots, b_m)\varphi = h_{\mathcal{B}}(b_1, \dots, b_m)\varphi = g_{\mathcal{C}}(b_1\varphi, \dots, b_m\varphi).$$

Since  $\varphi$  is surjective, this shows that  $\varphi$  is an epimorphism.

To prove (b), let  $n \geq 0$ ,  $\mathcal{A}_i = (A_i, \Sigma_i) \in \mathbf{K}$  for each  $i \in [n]$ ,  $\varkappa : \Omega \rightarrow \Sigma_1 \times \dots \times \Sigma_n$  be a mapping,  $\psi : \varkappa(\mathcal{A}_1, \dots, \mathcal{A}_n) \rightarrow \mathcal{B}$  be an isomorphism to an algebra  $\mathcal{B} = (B, \Omega)$ ,  $\mathcal{C} = (C, \Omega')$  be a g-subalgebra of  $\mathcal{B}$ , and  $(\iota, \varphi)$  a g-isomorphism from  $\mathcal{C}$  to  $\mathcal{D} = (D, \Gamma)$ . Then  $\mathcal{D}$  is a typical representative of  $S_g P_g(\mathbf{K})$ .

Let  $\lambda : \Gamma \rightarrow \Sigma_1 \times \dots \times \Sigma_n$  be the mapping such that  $\lambda(g) = \varkappa(\iota^{-1}(g))$  for each  $g \in \Gamma$ . Then  $\mathcal{E} = (C\psi^{-1}, \Gamma)$  is a subalgebra of the g-product  $\lambda(\mathcal{A}_1, \dots, \mathcal{A}_n) = (A_1 \times \dots \times A_n, \Gamma)$ . Indeed, let  $g \in \Gamma$ ,  $m \geq 0$ , and  $\mathbf{a}_1 = (a_{11}, \dots, a_{1n}), \dots, \mathbf{a}_m = (a_{m1}, \dots, a_{mn}) \in C\psi^{-1}$ . If  $\iota^{-1}(g) = h \in \Omega'$  and  $\varkappa(h) = (f_1, \dots, f_n)$ , then

$$\begin{aligned} g_{\lambda(\mathcal{A}_1, \dots, \mathcal{A}_n)}(\mathbf{a}_1, \dots, \mathbf{a}_m)\psi &= ((f_1)_{\mathcal{A}_1}(a_{11}, \dots, a_{m1}), \dots, (f_n)_{\mathcal{A}_n}(a_{1n}, \dots, a_{mn}))\psi \\ &= h_{\varkappa(\mathcal{A}_1, \dots, \mathcal{A}_n)}(\mathbf{a}_1, \dots, \mathbf{a}_m)\psi = h_{\mathcal{B}}(\mathbf{a}_1\psi, \dots, \mathbf{a}_m\psi) = h_{\mathcal{C}}(\mathbf{a}_1\psi, \dots, \mathbf{a}_m\psi) \end{aligned}$$

is in  $C$  since  $C$  is an  $\Omega'$ -closed subset of  $\mathcal{B}$ . To show that also  $\mathcal{D}$  is in  $SP_g(\mathbf{K})$ , we verify that  $\psi\varphi : \mathcal{E} \rightarrow \mathcal{D}$ , with  $\psi$  restricted to  $C\varphi^{-1}$ , is an isomorphism. For this, consider any  $g \in \Gamma$ ,  $m \geq 0$  and  $\mathbf{a}_1, \dots, \mathbf{a}_m \in C\psi^{-1}$ . If  $h \in \Omega'$  is the symbol such that  $\iota(h) = g$ , then

$$\begin{aligned} g\mathcal{E}(\mathbf{a}_1, \dots, \mathbf{a}_m)\psi\varphi &= g_{\lambda(A_1, \dots, A_n)}(\mathbf{a}_1, \dots, \mathbf{a}_m)\psi\varphi = h_C(\mathbf{a}_1\psi, \dots, \mathbf{a}_m\psi)\varphi \\ &= g_{\mathcal{D}}(\mathbf{a}_1\psi\varphi, \dots, \mathbf{a}_m\psi\varphi). \end{aligned}$$

Moreover, it is clear that  $\psi\varphi$  is bijective.  $\square$

Now we get the simplified representation of the  $V_g$ -operator.

**Proposition 4.12.**  $V_g = HSP_g$ .

*Proof.* Since  $\mathbf{K} \subseteq HSP_g(\mathbf{K}) \subseteq H_gS_gP_g(\mathbf{K}) = V_g(\mathbf{K})$  for any class  $\mathbf{K}$  of regular algebras, it suffices to verify that  $HSP_g(\mathbf{K})$  is a VRA. This holds because

$$\begin{aligned} S_g(HSP_g) &\leq HS_gSP_g \leq HS_gP_g \leq HSP_g, \\ H_g(HSP_g) &\leq H_gSP_g \leq HS_gP_g \leq HSP_g, \text{ and} \\ P_g(HSP_g) &\leq HP_gSP_g \leq HSP_gP_g = HSP_g, \end{aligned}$$

by Lemmas 4.9 and 4.11.  $\square$

Finally, let us note the following important fact.

**Lemma 4.13.** *Let  $\mathbf{K}$  be a VRA. If  $(\sigma, \theta)$  is a  $g$ -congruence of an unranked algebra  $\mathcal{A} = (A, \Sigma)$ , then  $\mathcal{A}/\theta \in \mathbf{K}$  iff  $\mathcal{A}/(\sigma, \theta) \in \mathbf{K}$ .*

*Proof.* It is easy to verify that  $(\sigma_{\mathfrak{t}}, 1_{A/\theta}) : \mathcal{A}/\theta \rightarrow \mathcal{A}/(\sigma, \theta)$  is a  $g$ -epimorphism. Hence,  $\mathcal{A}/\theta \in \mathbf{K}$  implies  $\mathcal{A}/(\sigma, \theta) \in \mathbf{K}$ .

Assume now that  $\mathcal{A}/(\sigma, \theta) \in \mathbf{K}$ . The  $g$ -derived algebra  $\sigma_{\mathfrak{t}}(\mathcal{A}/(\sigma, \theta))$  is actually the algebra  $\mathcal{A}/\theta$ . Indeed, both are  $\Sigma$ -algebras with the same set  $A/\theta$  of elements, and for any  $f \in \Sigma$ ,  $m \geq 0$  and  $a_1, \dots, a_m \in A$ ,

$$\begin{aligned} f_{\sigma_{\mathfrak{t}}(\mathcal{A}/(\sigma, \theta))}(a_1\theta, \dots, a_m\theta) &= (f\sigma)_{\mathcal{A}/(\sigma, \theta)}(a_1\theta, \dots, a_m\theta) = f_{\mathcal{A}}(a_1, \dots, a_m)\theta \\ &= f_{\mathcal{A}/\theta}(a_1\theta, \dots, a_m\theta). \end{aligned}$$

Hence,  $\mathcal{A}/\theta \in \mathbf{K}$  by Lemma 4.7.  $\square$

## 5 Regular congruences

We shall now introduce the congruences that play here the same role as the congruences of finite index in the theory of ranked varieties. In what follows,  $\mathcal{A} = (A, \Sigma)$  is again any unranked algebra. Let  $\text{FCon}(\mathcal{A}) := \{\theta \in \text{Con}(\mathcal{A}) \mid A/\theta \text{ finite}\}$  and let  $\text{FGCon}(\mathcal{A}) := \{(\sigma, \theta) \in \text{GCon}(\mathcal{A}) \mid \theta \in \text{FCon}(\mathcal{A})\}$ . If  $\theta \in \text{Eq}(A)$ , we treat  $A/\theta$  also as an alphabet and  $(A/\theta)^*$  as the free monoid generated by it.

**Definition 5.1.** We call a congruence  $\theta$  of  $\mathcal{A}$  *regular* if  $\theta \in \text{FCon}(\mathcal{A})$  and for every  $f \in \Sigma$  and every  $a \in A$ ,  $f_{\mathcal{A}/\theta}^{-1}(a\theta)$  is a regular language over  $A/\theta$ . A g-congruence  $(\sigma, \theta)$  of  $\mathcal{A}$  is *regular* if  $\theta$  is a regular congruence. Let  $\text{RCon}(\mathcal{A})$  and  $\text{RGCon}(\mathcal{A})$ , respectively, denote the sets of regular congruences and regular g-congruences of  $\mathcal{A}$ .  $\square$

For any  $\theta \in \text{Con}(\mathcal{A})$ , let  $\eta_\theta : A^* \rightarrow (A/\theta)^*$  be the monoid morphism such that  $a\eta_\theta = a\theta$  for each  $a \in A$ . It is easy to see that for all  $f \in \Sigma$ ,  $a \in A$  and  $w \in A^*$ ,  $w\eta_\theta \in f_{\mathcal{A}/\theta}^{-1}(a\theta)$  iff  $f_{\mathcal{A}}(w) \in a\theta$ , and hence  $f_{\mathcal{A}}^{-1}(a\theta) = f_{\mathcal{A}/\theta}^{-1}(a\theta)\eta_\theta^{-1}$ . Since  $\eta_\theta$  is surjective, this implies also  $f_{\mathcal{A}/\theta}^{-1}(a\theta) = f_{\mathcal{A}}^{-1}(a\theta)\eta_\theta$ . These equalities yield the following lemma that can be used for showing that a congruence is regular. Note that in  $f_{\mathcal{A}/\theta}^{-1}(a\theta)$ , the set  $a\theta$  is regarded as an element of  $A/\theta$ , but in  $f_{\mathcal{A}}^{-1}(a\theta)$  it is a subset of  $A$ .

**Lemma 5.2.** Let  $\theta \in \text{Con}(\mathcal{A})$ . For any  $f \in \Sigma$  and  $a \in A$ ,  $f_{\mathcal{A}/\theta}^{-1}(a\theta)$  is a regular language over  $A/\theta$  iff  $f_{\mathcal{A}}^{-1}(a\theta)$  is a regular language over  $A$ .  $\square$

Let us consider a simple example of a finite non-regular congruence.

**Example 5.3.** For  $\Sigma = \{f\}$  and  $X = \emptyset$ , let  $T$  be the  $\Sigma X$ -tree language such that

- (1)  $f \in T$ , and
- (2) for any  $n > 0$  and  $t_1, \dots, t_n \in T_\Sigma(X)$ ,  $f(t_1, \dots, t_n) \in T$  iff (a)  $n$  is even, (b)  $t_1, \dots, t_{n/2} \in T$ , and (c)  $t_{n/2+1}, \dots, t_n \notin T$ .

It is easy to see that  $\theta \in \text{FCon}(T_\Sigma(X))$  when  $\theta$  is defined by  $s\theta t : \Leftrightarrow (s \in T \leftrightarrow t \in T) \ (s, t \in T_\Sigma(X))$ . However,  $f_{T_\Sigma(X)/\theta}^{-1}(f\theta) = \{(f\theta)^n(f(f)\theta)^n \mid n \geq 0\}$  is not a regular language, and hence  $\theta$  is not a regular congruence.  $\square$

If  $\mathcal{A} = (A, \Sigma)$  is a regular algebra, then for any  $\theta \in \text{Con}(\mathcal{A})$ ,  $f \in \Sigma$  and  $a \in A$ ,  $f_{\mathcal{A}}^{-1}(a\theta)$  is the union of the finitely many regular sets  $f_{\mathcal{A}}^{-1}(b)$ , where  $b \in a\theta$ . Hence, Lemma 5.2 yields the following proposition.

**Proposition 5.4.** Every congruence of a regular algebra is regular.  $\square$

It is clear that  $\text{FCon}(\mathcal{A})$  and  $\text{FGCon}(\mathcal{A})$  are filters of the lattices  $\text{Con}(\mathcal{A})$  and  $\text{GCon}(\mathcal{A})$ , respectively. Moreover, the following hold.

**Lemma 5.5.** For any unranked algebra  $\mathcal{A} = (A, \Sigma)$ ,  $\text{RCon}(\mathcal{A})$  is a filter of the lattice  $\text{Con}(\mathcal{A})$ , and similarly,  $\text{RGCon}(\mathcal{A})$  is a filter of  $\text{GCon}(\mathcal{A})$ .

*Proof.* Since  $\text{RCon}(\mathcal{A})$  contains at least  $\nabla_A$ , it is nonempty.

If  $\theta, \rho \in \text{RCon}(\mathcal{A})$ , then clearly  $\theta \cap \rho \in \text{FCon}(\mathcal{A})$ . Moreover, for any  $f \in \Sigma$  and  $a \in A$ ,  $f_{\mathcal{A}}^{-1}(a(\theta \cap \rho)) = f_{\mathcal{A}}^{-1}(a\theta) \cap f_{\mathcal{A}}^{-1}(a\rho)$ , and hence also  $f_{\mathcal{A}/\theta \cap \rho}^{-1}(a(\theta \cap \rho))$  is regular by Lemma 5.2.

Next, let  $\theta \in \text{RCon}(\mathcal{A})$ ,  $\rho \in \text{Con}(\mathcal{A})$  and  $\theta \subseteq \rho$ . Of course,  $\rho \in \text{FCon}(\mathcal{A})$ . Moreover, for each  $a \in A$  there is a finite set of elements  $a_1, \dots, a_k \in A$  ( $k \geq 1$ )

such that  $a\rho = a_1\theta \cup \dots \cup a_k\theta$ , and hence  $f_{\mathcal{A}}^{-1}(a\rho) = f_{\mathcal{A}}^{-1}(a_1\theta) \cup \dots \cup f_{\mathcal{A}}^{-1}(a_k\theta)$  is a regular language for every  $f \in \Sigma$ . Hence  $\rho$  is regular by Lemma 5.2.

That  $\text{RGCon}(\mathcal{A})$  is a filter of  $\text{GCon}(\mathcal{A})$  follows immediately from the fact that  $\text{RCon}(\mathcal{A})$  is a filter of  $\text{Con}(\mathcal{A})$ .  $\square$

The following connection between regular algebras and regular congruences is a direct consequence of the definitions of these concepts.

**Proposition 5.6.** *If  $\theta$  is a congruence of an unranked algebra  $\mathcal{A}$ , then  $\mathcal{A}/\theta$  is a regular algebra exactly in case  $\theta$  is a regular congruence. Similarly, for any  $g$ -congruence  $(\sigma, \theta)$  of  $\mathcal{A}$ , the  $g$ -quotient  $\mathcal{A}/(\sigma, \theta)$  is regular if and only if  $(\sigma, \theta) \in \text{RGCon}(\mathcal{A})$ .*  $\square$

## 6 Syntactic congruences and algebras

The syntactic congruences and the syntactic algebras of unranked tree languages are defined in quite the same way as for ranked tree languages. Also here it is advantageous to define the notions for subsets of general unranked algebras.

**Definition 6.1.** The *syntactic congruence*  $\theta_H$  of a subset  $H \subseteq A$  of an unranked algebra  $\mathcal{A} = (A, \Sigma)$  is defined by

$$a\theta_H b :\Leftrightarrow (\forall p \in \text{Tr}(\mathcal{A}))(p(a) \in H \leftrightarrow p(b) \in H) \quad (a, b \in A),$$

and  $\text{SA}(H) := \mathcal{A}/\theta_H$  is the *syntactic algebra* of  $H$ . The natural morphism  $\varphi_H : \mathcal{A} \rightarrow \text{SA}(H)$ ,  $a \mapsto a\theta_H$ , is called the *syntactic morphism* of  $H$ .  $\square$

Let us recall that an equivalence  $\theta \in \text{Eq}(A)$  is said to *saturate* a subset  $H \subseteq A$  if  $H$  is the union of some  $\theta$ -classes. The following lemma can be proved exactly in the same way as for string languages or ordinary tree languages.

**Lemma 6.2.** *For any subset  $H \subseteq A$  of an unranked algebra  $\mathcal{A} = (A, \Sigma)$ ,  $\theta_H$  is the greatest congruence of  $\mathcal{A}$  that saturates  $H$ .*  $\square$

The following fact is an immediate consequence of Proposition 4.5.

**Proposition 6.3.** *If  $\mathcal{A} = (A, \Sigma)$  is an effectively given regular algebra, then the syntactic congruence  $\theta_H$  and the syntactic algebra  $\text{SA}(H)$  of any subset  $H \subseteq A$  can be effectively constructed.*  $\square$

**Definition 6.4.** An unranked algebra is called *syntactic* if it is isomorphic to the syntactic algebra of a subset of some unranked algebra. A subset  $D \subseteq A$  of an unranked algebra  $\mathcal{A} = (A, \Sigma)$  is *disjunctive* if  $\theta_D = \Delta_A$ .  $\square$

The following facts can be proved exactly as their well-known counterparts in the ranked case (cf. [23, 24, 26]).

**Proposition 6.5.** *An unranked algebra is syntactic iff it has a disjunctive subset.*  $\square$

**Proposition 6.6.** *Every finite gsd-irreducible unranked algebra is syntactic. Hence every VRA is generated by regular syntactic algebras.*  $\square$

It is easy to see that for any congruence  $\theta$  of an unranked algebra  $\mathcal{A} = (A, \Sigma)$ , there is a greatest equivalence  $M(\theta)$  on  $\Sigma$  such that  $(M(\theta), \theta) \in \text{GCon}(\mathcal{A})$ . We shall need the following properties of the  $M$ -operator (cf. Lemma 6.1 of [25]).

**Lemma 6.7.** *Let  $\mathcal{A} = (A, \Sigma)$  and  $\mathcal{B} = (B, \Omega)$  be unranked algebras.*

- (a) *If  $\theta, \theta' \in \text{Con}(\mathcal{A})$  and  $\theta \subseteq \theta'$ , then  $M(\theta) \subseteq M(\theta')$ .*
- (b) *For any set  $\{\theta_i \mid i \in I\}$  of congruences of  $\mathcal{A}$ ,  $M(\bigcap_{i \in I} \theta_i) = \bigcap_{i \in I} M(\theta_i)$ .*
- (c) *For any  $g$ -morphism  $(\iota, \varphi) : \mathcal{A} \rightarrow \mathcal{B}$  and any  $\theta \in \text{Con}(\mathcal{B})$ ,*

$$\iota \circ M(\theta) \circ \iota^{-1} \subseteq M(\varphi \circ \theta \circ \varphi^{-1}).$$

*If  $\varphi$  is surjective, then equality holds.*  $\square$

Although we shall mainly use the syntactic congruences and algebras defined above, the following input-reduced versions will be needed for the proof of our variety theorem.

**Definition 6.8.** The *reduced syntactic congruence* of a subset  $H$  of an unranked algebra  $\mathcal{A} = (A, \Sigma)$  is the  $g$ -congruence  $(\sigma_H, \theta_H)$  of  $\mathcal{A}$ , where  $\theta_H$  is the syntactic congruence of  $H$  and  $\sigma_H := M(\theta_H)$ , the *reduced syntactic algebra*  $\text{RA}(H)$  of  $H$  is the  $g$ -quotient  $\mathcal{A}/(\sigma_H, \theta_H) = (A/\theta_H, \Sigma/\sigma_H)$ , and the *syntactic  $g$ -morphism*  $(\iota_H, \varphi_H) : \mathcal{A} \rightarrow \text{RA}(H)$  is defined by  $\iota_H : f \mapsto f\sigma_H$  and  $\varphi_H : a \mapsto a\theta_H$ .  $\square$

The following proposition corresponds to Proposition 5.4 of [25].

**Proposition 6.9.** *Let  $\mathcal{A} = (A, \Sigma)$  and  $\mathcal{B} = (B, \Omega)$  be unranked algebras.*

- (a)  *$\theta_{A \setminus H} = \theta_H$  for every  $H \subseteq A$ .*
- (b)  *$\theta_H \cap \theta_K \subseteq \theta_{H \cap K}$  for all  $H, K \subseteq A$ .*
- (c)  *$\theta_H \subseteq \theta_{p^{-1}(H)}$  for all  $H \subseteq A$  and  $p \in \text{Tr}(\mathcal{A})$ .*
- (d) *If  $(\iota, \varphi) : \mathcal{A} \rightarrow \mathcal{B}$  is a  $g$ -morphism and  $H \subseteq B$ , then  $\varphi \circ \theta_H \circ \varphi^{-1} \subseteq \theta_{H\varphi^{-1}}$  and  $\iota \circ \sigma_H \circ \iota^{-1} \subseteq \sigma_{H\varphi^{-1}}$ , and equalities hold if  $(\iota, \varphi)$  is a  $g$ -epimorphism.*

*Proof.* Assertions (a)–(c) follow directly from the definition of syntactic congruences, so we prove just (d). For  $a, a' \in A$ ,

$$\begin{aligned} a\varphi \circ \theta_H \circ \varphi^{-1} a' &\Leftrightarrow (\forall q \in \text{Tr}(\mathcal{B})) (q(a\varphi) \in H \leftrightarrow q(a'\varphi) \in H) \\ &\Rightarrow (\forall p \in \text{Tr}(\mathcal{A})) (p_{\iota, \varphi}(a\varphi) \in H \leftrightarrow p_{\iota, \varphi}(a'\varphi) \in H) \\ &\Leftrightarrow (\forall p \in \text{Tr}(\mathcal{A})) (p(a) \in H\varphi^{-1} \leftrightarrow p(a') \in H\varphi^{-1}) \\ &\Leftrightarrow a\theta_{H\varphi^{-1}} a'. \end{aligned}$$

This proves the first inclusion of (d), and the second one now follows from Lemma 6.7:

$$\iota \circ \sigma_H \circ \iota^{-1} = \iota \circ M(\theta_H) \circ \iota^{-1} \subseteq M(\varphi \circ \theta_H \circ \varphi^{-1}) \subseteq M(\theta_{H\varphi^{-1}}) = \sigma_{H\varphi^{-1}}.$$

If  $(\iota, \varphi)$  is a g-epimorphism, the only “ $\Rightarrow$ ” in the proof of the first inclusion can be replaced by “ $\Leftrightarrow$ ”, and all inclusions become equalities.  $\square$

**Proposition 6.10.** *Let  $\mathcal{A} = (A, \Sigma)$  and  $\mathcal{B} = (B, \Omega)$  be unranked algebras.*

- (a)  $\text{SA}(A \setminus H) = \text{SA}(H)$  for every  $H \subseteq A$ .
- (b)  $\text{SA}(H \cap K) \preceq \text{SA}(H) \times \text{SA}(K)$  for all  $H, K \subseteq A$ .
- (c)  $\text{SA}(p^{-1}(H))$  is an epimorphic image of  $\text{SA}(H)$  for all  $H \subseteq A$  and  $p \in \text{Tr}(\mathcal{A})$ .
- (d) If  $(\iota, \varphi) : \mathcal{A} \rightarrow \mathcal{B}$  is a g-morphism and  $H \subseteq B$ , then  $\text{RA}(H\varphi^{-1}) \preceq_g \text{RA}(H)$ . If  $(\iota, \varphi)$  is a g-epimorphism, then  $\text{RA}(H\varphi^{-1}) \cong_g \text{RA}(H)$ .

*Proof.* Claims (a)–(c) follow directly from the corresponding parts of Proposition 6.9 (in the same way as for ranked algebras [24]).

To prove (d), assume first that  $(\iota, \varphi)$  is a g-epimorphism. It follows from Proposition 6.9(d) that the maps  $\psi : A/\theta_{H\varphi^{-1}} \rightarrow B/\theta_H$ ,  $a\theta_{H\varphi^{-1}} \mapsto (a\varphi)\theta_H$ , and  $\varkappa : \Sigma/\sigma_{H\varphi^{-1}} \rightarrow \Omega/\sigma_H$ ,  $f\sigma_{H\varphi^{-1}} \mapsto \iota(f)\sigma_H$ , are well-defined and injective. Clearly, they are also surjective, and for any  $f \in \Sigma$ ,  $m \geq 0$  and  $a_1, \dots, a_m \in A$ ,

$$\begin{aligned} (f\sigma_{H\varphi^{-1}})_{\text{RA}(H\varphi^{-1})}(a_1\theta_{H\varphi^{-1}}, \dots, a_m\theta_{H\varphi^{-1}})\psi &= (f_{\mathcal{A}}(a_1, \dots, a_m)\theta_{H\varphi^{-1}})\psi \\ &= (f_{\mathcal{A}}(a_1, \dots, a_m)\varphi)\theta_H = (\iota(f)_{\mathcal{B}}(a_1\varphi, \dots, a_m\varphi))\theta_H \\ &= (\iota(f)\sigma_H)_{\text{RA}(H)}((a_1\varphi)\theta_H, \dots, (a_m\varphi)\theta_H) \\ &= \varkappa(f\sigma_{H\varphi^{-1}})_{\text{RA}(H)}((a_1\theta_{H\varphi^{-1}})\psi, \dots, (a_m\theta_{H\varphi^{-1}})\psi), \end{aligned}$$

which shows that  $(\varkappa, \psi) : \text{RA}(H\varphi^{-1}) \rightarrow \text{RA}(H)$  is a g-isomorphism.

Consider now a general g-morphism  $(\iota, \varphi) : \mathcal{A} \rightarrow \mathcal{B}$ . Let  $\mathcal{C} = (C, \iota(\Sigma)/\sigma_H)$ , where  $C = A\varphi\varphi_H$  and  $\iota(\Sigma)/\sigma_H = \{\iota(f)\sigma_H \mid f \in \Sigma\}$ , be the image of  $\mathcal{A}$  in  $\text{RA}(H)$  under the g-morphism  $(\iota_H, \varphi\varphi_H) : \mathcal{A} \rightarrow \text{RA}(H)$ . The mappings

$$\varkappa : \Sigma \rightarrow \iota(\Sigma)/\sigma_H, f \mapsto \iota(f)\sigma_H, \text{ and } \psi : A \rightarrow C, a \mapsto (a\varphi)\theta_H,$$

define a g-epimorphism  $(\varkappa, \psi) : \mathcal{A} \rightarrow \mathcal{C}$ , and therefore  $\text{RA}(H\varphi^{-1}\psi\psi^{-1}) \cong_g \text{RA}(H\varphi^{-1}\psi)$  by the previous part of the proof. Also,  $\text{RA}(H\varphi^{-1}\psi) \preceq_g \text{RA}(H)$  because  $\text{RA}(H\varphi^{-1}\psi)$  is a g-image of the g-subalgebra  $\mathcal{C}$  of  $\text{RA}(H)$ . To obtain  $\text{RA}(H\varphi^{-1}) \preceq_g \text{RA}(H)$  it therefore suffices to show that  $H\varphi^{-1} = H\varphi^{-1}\psi\psi^{-1}$ . Of course,  $H\varphi^{-1} \subseteq H\varphi^{-1}\psi\psi^{-1}$ , and on the other hand,

$$\begin{aligned} H\varphi^{-1}\psi\psi^{-1} &= (H\varphi^{-1})\varphi \circ \varphi_H \circ (\varphi \circ \varphi_H)^{-1} = (H\varphi^{-1})\varphi \circ \theta_H \circ \varphi^{-1} \\ &\subseteq (H\varphi^{-1})\theta_{H\varphi^{-1}} = H\varphi^{-1}, \end{aligned}$$

where we used the facts that  $\psi = \varphi\varphi_H$ ,  $\varphi_H \circ \varphi_H^{-1} = \theta_H$ ,  $\varphi \circ \theta_H \circ \varphi^{-1} \subseteq \theta_{H\varphi^{-1}}$  and  $(H\varphi^{-1})\theta_{H\varphi^{-1}} = H\varphi^{-1}$ .  $\square$

Simple modifications of the proofs of statements (d) of Propositions 6.9 and 6.10 yield the following specializations of those statements.

**Corollary 6.11.** *Let  $\mathcal{A} = (A, \Sigma)$  and  $\mathcal{B} = (B, \Sigma)$  be unranked  $\Sigma$ -algebras. If  $\varphi : \mathcal{A} \rightarrow \mathcal{B}$  is a morphism and  $H \subseteq B$ , then  $\varphi \circ \theta_H \circ \varphi^{-1} \subseteq \theta_{H\varphi^{-1}}$  and  $\text{SA}(H\varphi^{-1}) \preceq \text{SA}(H)$ . If  $\varphi$  is an epimorphism, then  $\varphi \circ \theta_H \circ \varphi^{-1} = \theta_{H\varphi^{-1}}$  and  $\text{SA}(H\varphi^{-1}) \cong \text{SA}(H)$ .  $\square$*

The *syntactic congruence*  $\theta_T$ , the *syntactic algebra*  $\text{SA}(T)$ , the *syntactic morphism*  $\varphi_T : \mathcal{T}_\Sigma(X) \rightarrow \text{SA}(T)$ , the *reduced syntactic congruence*  $(\sigma_T, \theta_T)$  and the *reduced syntactic algebra*  $\text{RA}(T)$  as well as the *syntactic  $g$ -morphism*  $(\iota_T, \varphi_T) : \mathcal{T}_\Sigma(X) \rightarrow \text{RA}(T)$  of a  $\Sigma X$ -tree language  $T$  are defined by regarding  $T$  as a subset of the term algebra  $\mathcal{T}_\Sigma(X)$ . Since the translations of  $\mathcal{T}_\Sigma(X)$  are given by  $\Sigma X$ -contexts, we have

$$s \theta_T t \Leftrightarrow (\forall p \in C_\Sigma(X))(p(s) \in T \leftrightarrow p(t) \in T),$$

for all  $s, t \in \mathcal{T}_\Sigma(X)$ . Let us note that the syntactic congruences of unranked tree languages appear in [6] under the name “top congruence”.

## 7 Recognizable unranked tree languages

Within our framework it is natural to define recognizability as follows.

**Definition 7.1.** An unranked  $\Sigma$ -algebra  $\mathcal{A} = (A, \Sigma)$  *recognizes* an unranked  $\Sigma X$ -tree language  $T$  if there exist a morphism  $\varphi : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{A}$  and a subset  $F \subseteq A$  such that  $T = F\varphi^{-1}$ , and we call  $T$  *recognizable* if it is recognized by a regular  $\Sigma$ -algebra. The set of all recognizable unranked  $\Sigma X$ -tree languages is denoted by  $\text{Rec}(\Sigma, X)$ .  $\square$

Exactly as in the ranked case, the syntactic algebra of any given unranked tree language  $T$  is in a definite sense the minimal algebra recognizing  $T$ .

**Proposition 7.2.** *A  $\Sigma$ -algebra  $\mathcal{A}$  recognizes an unranked  $\Sigma X$ -tree language  $T$  if and only if  $\text{SA}(T) \preceq \mathcal{A}$ .*

*Proof.* It is clear, quite generally, that any tree language recognized by a sub-algebra or an epimorphic image of an algebra  $\mathcal{A}$ , is recognized by  $\mathcal{A}$ , too. Since  $\text{SA}(T)$  recognizes  $T$  (as  $T = T\varphi_T\varphi_T^{-1}$ ), this means that if  $\text{SA}(T) \preceq \mathcal{A}$ , then also  $\mathcal{A}$  recognizes  $T$ . The converse holds by Corollary 6.11: if there exist a morphism  $\varphi : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{A}$  and a subset  $F$  of  $\mathcal{A}$  such that  $T = F\varphi^{-1}$ , then  $\text{SA}(T) = \text{SA}(F\varphi^{-1}) \preceq \text{SA}(F) \preceq \mathcal{A}$ .  $\square$

It should be obvious that the above notion of recognizability is the same as the one arrived at via deterministic or nondeterministic automata on unranked trees as defined in [9, 16, 22], for example. Indeed, for any regular algebra  $\mathcal{A} = (A, \Sigma)$ , we can introduce finite (“horizontal”) automata to compute the operations  $f_{\mathcal{A}}$  and, conversely, the functions computed by the horizontal automata of a deterministic automaton on unranked trees are the operations of a

regular algebra. However, since we are not directly concerned with computational aspects, the algebraic definition is more convenient here.

Next we give a Myhill-Nerode theorem that characterizes the recognizability of unranked tree languages in terms of regular congruences.

**Proposition 7.3.** *For any unranked tree language  $T \subseteq T_\Sigma(X)$ , the following statements are equivalent:*

- (a)  $T \in \text{Rec}(\Sigma, X)$ ;
- (b)  $T$  is saturated by a regular congruence of  $\mathcal{T}_\Sigma(X)$ ;
- (c) the syntactic congruence  $\theta_T$  is regular.

*Proof.* Let us first prove the equivalence of (a) and (b). If  $T \in \text{Rec}(\Sigma, X)$ , then there exist a regular algebra  $\mathcal{A} = (A, \Sigma)$ , a morphism  $\varphi : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{A}$  and subset  $F \subseteq A$  such that  $T = F\varphi^{-1}$ . It is clear that  $T$  is saturated by  $\theta := \ker \varphi$ . Moreover, we may assume that  $\varphi$  is surjective. Then  $\mathcal{T}_\Sigma(X)/\theta \cong \mathcal{A}$  and hence  $\theta \in \text{RCon}(\mathcal{T}_\Sigma(X))$  by Proposition 5.6. On the other hand, if  $T$  is saturated by a regular congruence  $\theta \in \text{RCon}(\mathcal{T}_\Sigma(X))$ , then  $T = T\theta_{\mathfrak{t}}\theta_{\mathfrak{t}}^{-1}$  means that  $T$  is recognized by the regular algebra  $\mathcal{T}_\Sigma(X)/\theta$ .

If  $T$  is saturated by a congruence  $\theta \in \text{RCon}(\mathcal{T}_\Sigma(X))$ , then  $\theta \subseteq \theta_T$  by Lemma 6.2, and hence  $\theta_T$  is regular by Lemma 5.5. Therefore, (b) implies (c), and the converse holds by Lemma 6.2.  $\square$

Let us note that in [6] it was stated (as Lemma 8.2), in different terms, that if  $T \in \text{Rec}(\Sigma, X)$ , then  $\theta_T$  is of finite index, but the example meant to show that the converse does not hold, appears incorrect. Nevertheless, their Theorem 1 essentially expresses the equivalence of (a) and (c) of our Proposition 7.3; the condition concerning “local views” seems to be a somewhat intricate way to define the regularity of a congruence.

The following corollary is an immediate consequence of Proposition 7.3.

**Corollary 7.4.** *An unranked tree language  $T$  is recognizable if and only if the syntactic algebra  $\text{SA}(T)$  is regular.*  $\square$

We shall now note that the family of recognizable unranked tree languages is closed under the operations that will define our varieties of unranked tree languages.

**Proposition 7.5.** *The following statements hold for all operator alphabets  $\Sigma$  and  $\Omega$  and all leaf alphabets  $X$  and  $Y$ .*

- (a)  $\emptyset \in \text{Rec}(\Sigma, X)$ , and  $\text{Rec}(\Sigma, X)$  is closed under all Boolean operations.
- (b) If  $T \in \text{Rec}(\Sigma, X)$ , then  $p^{-1}(T) := \{t \in T_\Sigma(X) \mid p(t) \in T\} \in \text{Rec}(\Sigma, X)$  for every context  $p \in C_\Sigma(X)$ .
- (c) If  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{T}_\Omega(Y)$  is a  $g$ -morphism, then  $T\varphi^{-1} \in \text{Rec}(\Sigma, X)$  for every  $T \in \text{Rec}(\Omega, Y)$ .



*Proof.* Clearly,  $\emptyset$  and  $T_\Sigma(X)$  are recognized by any  $\Sigma$ -algebra, and the rest of the proposition follows from Corollary 7.4 and Propositions 4.8 and 6.10.  $\square$

The sets  $p^{-1}(T)$  play an important role in the variety theory and we shall need the following fact.

**Lemma 7.6.** *If  $T \in \text{Rec}(\Sigma, X)$ , then the set  $\{p^{-1}(T) \mid p \in C_\Sigma(X)\}$  is finite.*

*Proof.* By Proposition 6.9(c) every set  $p^{-1}(T)$  is saturated by  $\theta_T$ . On the other hand, it follows from Proposition 7.3 that  $\theta_T$  has just finitely many equivalence classes. Hence, the number of different sets  $p^{-1}(T)$  must be finite, too.  $\square$

Let us say that a recognizable unranked  $\Sigma X$ -tree language  $T$  is *effectively given* if  $T = F\varphi^{-1}$ , where  $\varphi : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{A}$  is an effectively given morphism from  $\mathcal{T}_\Sigma(X)$  to an effectively given regular algebra  $\mathcal{A} = (A, \Sigma)$  and  $F \subseteq A$  is also effectively given.

**Proposition 7.7.** *If  $T \in \text{Rec}(\Sigma, X)$  is effectively given, then  $\text{SA}(T)$  can be effectively constructed.*

*Proof.* Assume that  $T = F\varphi^{-1}$ , where the morphism  $\varphi : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{A}$ , the regular algebra  $\mathcal{A} = (A, \Sigma)$  and the subset  $F \subseteq A$  are effectively given. Obviously, we may assume that  $\varphi$  is an epimorphism. Then  $\text{SA}(T) \cong \text{SA}(F)$  by Proposition 6.10, and  $\text{SA}(F)$  can be constructed by Proposition 6.3.  $\square$

## 8 Varieties of unranked tree languages

A *family of (recognizable) unranked tree languages* is a mapping  $\mathcal{V}$  that assigns to each pair  $\Sigma, X$  a set  $\mathcal{V}(\Sigma, X)$  of (recognizable)  $\Sigma X$ -tree languages. We write  $\mathcal{V} = \{\mathcal{V}(\Sigma, X)\}$  with the understanding that  $\Sigma$  and  $X$  range over all operator alphabets and leaf alphabets, respectively. The inclusion relation, unions and intersections of these families are defined by the natural componentwise conditions. In particular, if  $\mathcal{U} = \{\mathcal{U}(\Sigma, X)\}$  and  $\mathcal{V} = \{\mathcal{V}(\Sigma, X)\}$  are two such families, then  $\mathcal{U} \subseteq \mathcal{V}$  means that  $\mathcal{U}(\Sigma, X) \subseteq \mathcal{V}(\Sigma, X)$  for all  $\Sigma$  and  $X$ , and  $\mathcal{U} \cap \mathcal{V} = \{\mathcal{U}(\Sigma, X) \cap \mathcal{V}(\Sigma, X)\}$ .

**Definition 8.1.** A *variety of unranked tree languages (VUT)* is a family of recognizable unranked tree languages  $\mathcal{V} = \{\mathcal{V}(\Sigma, X)\}$  for which the following hold for all  $\Sigma, \Omega, X$  and  $Y$ .

- (V1)  $\emptyset \neq \mathcal{V}(\Sigma, X) \subseteq \text{Rec}(\Sigma, X)$ .
- (V2) If  $T \in \mathcal{V}(\Sigma, X)$ , then also  $T_\Sigma(X) \setminus T$  belongs to  $\mathcal{V}(\Sigma, X)$ .
- (V3) If  $T, U \in \mathcal{V}(\Sigma, X)$ , then  $T \cap U \in \mathcal{V}(\Sigma, X)$ .
- (V4) If  $T \in \mathcal{V}(\Sigma, X)$ , then  $p^{-1}(T) \in \mathcal{V}(\Sigma, X)$  for every  $p \in C_\Sigma(X)$ .
- (V5) If  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{T}_\Omega(Y)$  is a g-morphism, then  $T\varphi^{-1} \in \mathcal{V}(\Sigma, X)$  for every  $T \in \mathcal{V}(\Omega, Y)$ .

Let **VUT** denote the class of all VUTs.  $\square$

It is obvious that the intersection of any family of VUTs is a VUT. Hence,  $(\mathbf{VUT}, \subseteq)$  is a complete lattice. It is also clear that the union of any directed family of VUTs is a VUT. The least VUT is  $Triv := \{\{\emptyset, T_\Sigma(X)\}\}$  and the greatest one is the family  $Rec := \{Rec(\Sigma, X)\}$  of all recognizable unranked tree languages. The following fact will be used in the proof of the variety theorem.

**Proposition 8.2.** *If  $\mathcal{V} = \{\mathcal{V}(\Sigma, X)\}$  is a VUT and  $T \in \mathcal{V}(\Sigma, X)$  for some  $\Sigma$  and  $X$ , then every  $\theta_T$ -class is also in  $\mathcal{V}(\Sigma, X)$ .*

*Proof.* It follows from the definition of  $\theta_T$  that for any  $t \in T_\Sigma(X)$ ,

$$t\theta_T = \bigcap \{p^{-1}(T) \mid p \in C_\Sigma(X), p(t) \in T\} \setminus \bigcup \{p^{-1}(T) \mid p \in C_\Sigma(X), p(t) \notin T\}.$$

By Lemma 7.6, this shows that  $t\theta_T$  is in  $\mathcal{V}(\Sigma, X)$ .  $\square$

As in the ranked case, many VUTs have natural definitions based on congruences of term algebras. Hence, before considering further examples of VUTs, we introduce the systems of congruences that yield varieties of unranked tree languages. For any mapping  $\varphi : A \rightarrow B$  and any  $\theta \in \text{Eq}(B)$ , let  $\theta_\varphi := \varphi \circ \theta \circ \varphi^{-1}$ . Then  $\theta_\varphi \in \text{Eq}(A)$ , and if  $B/\theta$  is finite, then so is  $A/\theta_\varphi$ .

**Lemma 8.3.** *If  $(\omega, \theta) \in \text{RGCon}(\mathcal{T}_\Omega(Y))$ , then  $(\omega_\iota, \theta_\varphi) \in \text{RGCon}(\mathcal{T}_\Sigma(X))$  for every  $g$ -morphism  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{T}_\Omega(Y)$ .*

*Proof.* If  $f, g \in \Sigma$ ,  $s_1, \dots, s_m, t_1, \dots, t_m \in T_\Sigma(X)$  ( $m \geq 0$ ) are such that  $f\omega_\iota g$  and  $s_i\theta_\varphi t_i$  for every  $i \in [m]$ , then  $\iota(f)\omega\iota(g)$  and  $s_i\varphi\theta t_i\varphi$  for every  $i \in [m]$ , and therefore

$$f(s_1, \dots, s_m)\varphi = \iota(f)(s_1\varphi, \dots, s_m\varphi) \equiv_\theta \iota(g)(t_1\varphi, \dots, t_m\varphi) = g(t_1, \dots, t_m)\varphi,$$

which shows that  $f\mathcal{T}_\Sigma(X)(s_1, \dots, s_m)\theta_\varphi g\mathcal{T}_\Sigma(X)(t_1, \dots, t_m)$  as required.  $\square$

By a *family of regular  $g$ -congruences* we mean a mapping  $\mathcal{C}$  that assigns to each pair  $\Sigma, X$  a subset  $\mathcal{C}(\Sigma, X)$  of  $\text{RGCon}(\mathcal{T}_\Sigma(X))$ . Again, we write  $\mathcal{C} = \{\mathcal{C}(\Sigma, X)\}$  and order these families by the componentwise inclusion relation.

**Definition 8.4.** A family of regular  $g$ -congruences  $\mathcal{C} = \{\mathcal{C}(\Sigma, X)\}$  is a *variety of regular  $g$ -congruences (VRC)* if the following three conditions hold for all  $\Sigma, \Omega, X$  and  $Y$ .

- (C1) For every  $\sigma \in \text{Eq}(\Sigma)$ ,  $\mathcal{C}(\Sigma, X)_\sigma := \{\theta \in \text{RCon}(\mathcal{T}_\Sigma(X)) \mid (\sigma, \theta) \in \mathcal{C}(\Sigma, X)\}$  is a filter of  $\text{RCon}(\mathcal{T}_\Sigma(X))$ .
- (C2) If  $(\sigma, \theta) \in \mathcal{C}(\Sigma, X)$  and  $(\sigma', \theta) \in \text{RGCon}(\mathcal{T}_\Sigma(X))$ , then  $(\sigma', \theta) \in \mathcal{C}(\Sigma, X)$ .
- (C3) If  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{T}_\Omega(Y)$  is any  $g$ -morphism, then  $(\omega_\iota, \theta_\varphi) \in \mathcal{C}(\Sigma, X)$  for every  $(\omega, \theta) \in \mathcal{C}(\Omega, Y)$ .  $\square$

We shall now show that any variety of regular congruences yields a variety of unranked tree languages. For any family  $\mathcal{C} = \{\mathcal{C}(\Sigma, X)\}$  of regular g-congruences, let  $\mathcal{C}^t$  be the family of recognizable unranked tree languages such that for all  $\Sigma$  and  $X$ ,

$$\mathcal{C}^t(\Sigma, X) := \{T \subseteq T_\Sigma(X) \mid (\Delta_\Sigma, \theta_T) \in \mathcal{C}(\Sigma, X)\}.$$

**Proposition 8.5.** *If  $\mathcal{C} = \{\mathcal{C}(\Sigma, X)\}$  is a VRC, then  $\mathcal{C}^t = \{\mathcal{C}^t(\Sigma, X)\}$  is a VUT.*

*Proof.* Most of the proposition follows directly from the definitions involved and Proposition 6.9. Let us verify conditions (V1) and (V5) of Definition 8.1.

Firstly, for any  $\Sigma$  and  $X$ ,  $\mathcal{C}^t(\Sigma, X) \neq \emptyset$  because  $(\Delta_\Sigma, \nabla_{T_\Sigma(X)})$  certainly is in  $\mathcal{C}(\Sigma, X)$  and  $\theta_\emptyset = \nabla_{T_\Sigma(X)}$ . If  $T \in \mathcal{C}^t(\Sigma, X)$ , then  $(\Delta_\Sigma, \theta_T) \in \mathcal{C}(\Sigma, X)$  and hence  $\theta_T \in \text{RCon}(\mathcal{T}_\Sigma(X))$ , and by Proposition 7.3 this means that  $T$  is recognizable. Hence,  $\mathcal{C}^t$  satisfies (V1).

If  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{T}_\Omega(Y)$  is any g-morphism and  $T \in \mathcal{C}^t(\Omega, Y)$ , then we have  $(\Delta_\Omega, \theta_T) \in \mathcal{C}(\Omega, Y)$  which, by condition (C3), implies

$$(\iota \circ \Delta_\Omega \circ \iota^{-1}, \varphi \circ \theta_T \circ \varphi^{-1}) \in \mathcal{C}(\Sigma, X).$$

Furthermore,  $(\Delta_\Sigma, \varphi \circ \theta_T \circ \varphi^{-1}) \in \text{RGCon}(\mathcal{T}_\Sigma(X))$  because  $\theta_T \in \text{RCon}(\mathcal{T}_\Omega(Y))$  implies that  $\varphi \circ \theta_T \circ \varphi^{-1} \in \text{RCon}(\mathcal{T}_\Sigma(X))$ . Hence,  $(\Delta_\Sigma, \varphi \circ \theta_T \circ \varphi^{-1}) \in \mathcal{C}(\Sigma, X)$  by condition (C2). On the other hand,  $\varphi \circ \theta_T \circ \varphi^{-1} \subseteq \theta_{T\varphi^{-1}}$  by Proposition 6.9(d), and therefore  $(\Delta_\Sigma, \theta_{T\varphi^{-1}}) \in \mathcal{C}(\Sigma, X)$  by (C1). This means that  $T\varphi^{-1} \in \mathcal{C}^t(\Sigma, X)$  and therefore  $\mathcal{C}^t$  satisfies (V5).  $\square$

In many of the following examples, we define a whole family  $\mathcal{F}$  of VUTs indexed by some parameter(s). If  $\mathcal{F}$  forms an ascending chain or, more generally, is directed, then the union  $\bigcup \mathcal{F}$  is always also a VUT. In the ranked case the basic varieties of tree languages forming such a family  $\mathcal{F}$  are usually defined by so-called *principal varieties of congruences* [24, 26] that consist of principal filters of the term algebras. Here the same purpose will be served by the following more general notion.

**Definition 8.6.** For each pair  $\Sigma, X$ , let  $\theta(\Sigma, X)$  be a congruence of  $\mathcal{T}_\Sigma(X)$ . We call  $\Theta = \{\theta(\Sigma, X)\}_{\Sigma, X}$  a *consistent system of congruences* if  $\theta(\Sigma, X) \subseteq \varphi \circ \theta(\Omega, Y) \circ \varphi^{-1}$  for all alphabets  $\Sigma, \Omega, X$  and  $Y$ , and every g-morphism  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{T}_\Omega(Y)$ .

For any such system of congruences  $\Theta = \{\theta(\Sigma, X)\}_{\Sigma, X}$ , and all  $\Sigma$  and  $X$ , let

$$\mathcal{C}_\Theta(\Sigma, X) := \{(\sigma, \theta) \in \text{RGCon}(\mathcal{T}_\Sigma(X)) \mid \theta(\Sigma, X) \subseteq \theta\},$$

and let  $\mathcal{C}_\Theta := \{\mathcal{C}_\Theta(\Sigma, X)\}$  be the thus defined family of regular congruences.  $\square$

**Lemma 8.7.** *For any consistent system of congruences  $\Theta$ ,  $\mathcal{C}_\Theta$  is a VRC.*

*Proof.* (C1) If  $\sigma \in \text{Eq}(\Sigma)$ , then  $(\sigma, \nabla_{T_\Sigma(X)}) \in \mathcal{C}_\Theta(\Sigma, X)$ , and hence  $\mathcal{C}_\Theta(\Sigma, X)_\sigma \neq \emptyset$ . If  $\theta \subseteq \rho$  and  $\theta \in \mathcal{C}_\Theta(\Sigma, X)_\sigma$ , then  $\theta(\Sigma, X) \subseteq \theta \subseteq \rho$ . On the other hand,  $(\sigma, \theta) \in \text{RGCon}(\mathcal{T}_\Sigma(X))$  implies  $(\sigma, \rho) \in \text{RGCon}(\mathcal{T}_\Sigma(X))$  by Lemma 5.5. Hence

$\rho \in \mathcal{C}_\Theta(\Sigma, X)_\sigma$ . If  $\theta, \rho \in \mathcal{C}_\Theta(\Sigma, X)_\sigma$ , then  $\theta(\Sigma, X) \subseteq \theta, \rho$  and  $(\sigma, \theta), (\sigma, \rho) \in \text{RGCon}(\mathcal{T}_\Sigma(X))$ , and therefore  $\theta(\Sigma, X) \subseteq \theta \cap \rho$  and – again by Lemma 5.5,  $(\sigma, \theta \cap \rho) = (\sigma, \theta) \wedge (\sigma, \rho) \in \text{RGCon}(\mathcal{T}_\Sigma(X))$ . This means that  $\theta \cap \rho \in \mathcal{C}_\Theta(\Sigma, X)_\sigma$ , and thus we have shown that  $\mathcal{C}_\Theta(\Sigma, X)_\sigma$  is a filter in  $\text{RCon}(\mathcal{T}_\Sigma(X))$ .

(C2) If  $(\sigma, \theta) \in \mathcal{C}_\Theta(\Sigma, X)$  and  $(\sigma', \theta) \in \text{RGCon}(\mathcal{T}_\Sigma(X))$ , then also  $(\sigma', \theta) \in \mathcal{C}_\Theta(\Sigma, X)$  because  $\theta(\Sigma, X) \subseteq \theta$  by the first assumption.

(C3) If  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{T}_\Omega(Y)$  is a g-morphism and  $(\omega, \theta) \in \mathcal{C}_\Theta(\Omega, Y)$ , then  $(\omega_\iota, \theta_\varphi) \in \text{RGCon}(\mathcal{T}_\Sigma(X))$  by Lemma 8.3, and  $\theta(\Sigma, X) \subseteq \theta(\Omega, Y)_\varphi \subseteq \theta_\varphi$  by our assumption about the  $\theta(\Sigma, X)$ -congruences and the fact that  $\theta(\Omega, Y) \subseteq \theta$ . Hence  $(\omega_\iota, \theta_\varphi) \in \mathcal{C}_\Theta(\Sigma, X)$ .  $\square$

Let us call a VRC *quasi-principal* if it is defined this way by a consistent system of congruences  $\Theta$ . The corresponding VUT  $\mathcal{C}_\Theta^t$  is written as  $\mathcal{V}_\Theta = \{\mathcal{V}_\Theta(\Sigma, X)\}$  and also it is said to be *quasi-principal*. The following description of  $\mathcal{V}_\Theta$  is a direct consequence of its definition.

**Lemma 8.8.** *Let  $\Theta = \{\theta(\Sigma, X)\}_{\Sigma, X}$  be any consistent system of congruences. Then  $\mathcal{V}_\Theta(\Sigma, X) = \{T \in \text{Rec}(\Sigma, X) \mid \theta(\Sigma, X) \subseteq \theta_T\}$  for all  $\Sigma$  and  $X$ .*  $\square$

## 9 The variety theorem

We shall now prove that the following maps  $\mathbf{K} \mapsto \mathbf{K}^t$  and  $\mathcal{V} \mapsto \mathcal{V}^a$  form a pair of mutually inverse isomorphisms between the lattices  $(\mathbf{VRA}, \subseteq)$  and  $(\mathbf{VUT}, \subseteq)$ .

**Definition 9.1.** For any VRA  $\mathbf{K}$ , let  $\mathbf{K}^t = \{\mathbf{K}^t(\Sigma, X)\}$  be the family of recognizable unranked tree languages in which, for all  $\Sigma$  and  $X$ ,

$$\mathbf{K}^t(\Sigma, X) := \{T \subseteq T_\Sigma(X) \mid \text{SA}(T) \in \mathbf{K}\}.$$

For any VUT  $\mathcal{V} = \{\mathcal{V}(\Sigma, X)\}$ , let  $\mathcal{V}^a$  be the VRA generated by the algebras  $\text{SA}(T)$ , where  $T \in \mathcal{V}(\Sigma, X)$  for some  $\Sigma$  and  $X$ .  $\square$

Note that  $\mathcal{V}^a$  is a well-defined VRA for every VUT  $\mathcal{V}$  because all the algebras  $\text{SA}(T)$  with  $T \in \mathcal{V}(\Sigma, X)$  are regular. Note also that the maps  $\mathbf{K} \mapsto \mathbf{K}^t$  and  $\mathcal{V} \mapsto \mathcal{V}^a$  were defined in terms of syntactic algebras, but it follows from Lemma 4.13 that definitions that use reduced syntactic algebras (similarly as in [25]) would give the same maps.

**Lemma 9.2.** *For any VRA  $\mathbf{K}$ ,  $\mathbf{K}^t$  is a VUT.*

*Proof.* It follows from Corollary 7.4 that  $\mathbf{K}^t(\Sigma, X) \subseteq \text{Rec}(\Sigma, X)$  for all  $\Sigma$  and  $X$ . Moreover,  $\mathbf{K}^t(\Sigma, X) \neq \emptyset$  because  $\mathbf{K}$  contains at least the trivial  $\Sigma$ -algebras. Hence,  $\mathbf{K}^t$  satisfies condition (V1) of Definition 8.1. Conditions (V2)–(V4) follow immediately from Proposition 6.10 and the fact that  $\mathbf{K}$  is a VRA. As to (V5), we may argue as follows. If  $T \in \mathbf{K}^t(\Omega, Y)$ , then  $\text{SA}(T) \in \mathbf{K}$ . By Lemma 4.13 this implies  $\text{RA}(T) \in \mathbf{K}$  which by Proposition 6.10(d) implies that  $\text{RA}(T\varphi^{-1}) \in \mathbf{K}$ . Using again Lemma 4.13, we get  $\text{SA}(T\varphi^{-1}) \in \mathbf{K}$  from which  $T\varphi^{-1} \in \mathbf{K}^t(\Sigma, X)$  follows.  $\square$

It is clear that the maps  $\mathbf{K} \mapsto \mathbf{K}^t$  and  $\mathcal{V} \mapsto \mathcal{V}^a$  are isotone. To prove that they define isomorphisms between the lattices  $(\mathbf{VRA}, \subseteq)$  and  $(\mathbf{VUT}, \subseteq)$  it therefore suffices to show that they are inverses of each other.

**Lemma 9.3.**  $\mathbf{K}^{ta} = \mathbf{K}$  for every VRA  $\mathbf{K}$ .

*Proof.* The VRA  $\mathbf{K}^{ta}$  is generated by the algebras  $\text{SA}(T)$ , where  $T \in \mathbf{K}^t(\Sigma, X)$  for some  $\Sigma$  and  $X$ , but these algebras are, by the definition of  $\mathbf{K}^t$ , also in  $\mathbf{K}$ . Hence,  $\mathbf{K}^{ta} \subseteq \mathbf{K}$ .

On the other hand, by Proposition 6.6,  $\mathbf{K}$  is generated by regular syntactic algebras. Let  $\mathcal{A} = (A, \Sigma)$  be any such generating algebra. If  $X$  is a sufficiently large leaf alphabet, there is an epimorphism  $\varphi : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{A}$ . Furthermore,  $\mathcal{A}$  has a disjunctive subset  $D$  by Proposition 6.5. The  $\Sigma X$ -tree language  $T := D\varphi^{-1}$  is recognizable, and  $\text{SA}(T) \cong \text{SA}(D)$  by Corollary 6.11. On the other hand,  $\mathcal{A} \cong \text{SA}(D)$  because  $D$  disjunctive, and therefore also  $\text{SA}(T) \in \mathbf{K}$ , which shows that  $T \in \mathbf{K}^t(\Sigma, X)$ . As this means that  $\text{SA}(T) \in \mathbf{K}^{ta}$ , we also get  $\mathcal{A} \in \mathbf{K}^{ta}$  and we can conclude that  $\mathbf{K} \subseteq \mathbf{K}^{ta}$ .  $\square$

**Lemma 9.4.**  $\mathcal{V}^{at} = \mathcal{V}$  for every VUT  $\mathcal{V}$ .

*Proof.* If  $T \in \mathcal{V}(\Sigma, X)$ , then  $\text{SA}(T) \in \mathcal{V}^a$  implies  $T \in \mathcal{V}^{at}(\Sigma, X)$ , and hence  $\mathcal{V} \subseteq \mathcal{V}^{at}$ .

If  $T \in \mathcal{V}^{at}(\Sigma, X)$ , then  $\text{SA}(T) \in \mathcal{V}^a$  and by Proposition 4.12 this means that

$$\text{SA}(T) \preceq \varkappa(\text{SA}(U_1), \dots, \text{SA}(U_n)),$$

where  $n \geq 0$ ,  $U_1 \in \mathcal{V}(\Sigma_1, X_1)$ ,  $\dots$ ,  $U_n \in \mathcal{V}(\Sigma_n, X_n)$  for some alphabets  $\Sigma_1, \dots, \Sigma_n$  and  $X_1, \dots, X_n$ , and  $\varkappa$  is a mapping from  $\Sigma$  to  $\Sigma_1 \times \dots \times \Sigma_n$ .

For each  $i \in [n]$ , denote  $\mathcal{T}_{\Sigma_i}(X_i)$  by  $\mathcal{T}_i$ , and let  $\text{SA}(U_i) = (A_i, \Sigma_i)$ . Furthermore, let  $\varphi_i : \mathcal{T}_i \rightarrow \text{SA}(U_i)$ ,  $t \mapsto t\theta_{U_i}$ , be the syntactic morphism of  $U_i$ . By Proposition 7.2, there exist a morphism

$$\varphi : \mathcal{T}_\Sigma(X) \rightarrow \varkappa(\text{SA}(U_1), \dots, \text{SA}(U_n))$$

and a subset  $F \subseteq A_1 \times \dots \times A_n$  such that  $T = F\varphi^{-1}$ . For each  $i \in [n]$ , define  $\lambda_i : \Sigma \rightarrow \Sigma_i$  so that for any  $f \in \Sigma$ , if  $\varkappa(f) = (f_1, \dots, f_n)$ , then  $\lambda_i(f) = f_i$ . The syntactic morphisms  $\varphi_i$  yield an epimorphism

$$\eta : \varkappa(\mathcal{T}_1, \dots, \mathcal{T}_n) \rightarrow \varkappa(\text{SA}(U_1), \dots, \text{SA}(U_n)), (t_1, \dots, t_n) \mapsto (t_1\varphi_1, \dots, t_n\varphi_n),$$

and for each  $i \in [n]$ , we get the g-morphisms

$$(\lambda_i, \tau_i) : \varkappa(\mathcal{T}_1, \dots, \mathcal{T}_n) \rightarrow \mathcal{T}_i \text{ and } (\lambda_i, \pi_i) : \varkappa(\text{SA}(U_1), \dots, \text{SA}(U_n)) \rightarrow \text{SA}(U_i),$$

where  $\tau_i : (t_1, \dots, t_n) \mapsto t_i$  and  $\pi_i : (a_1, \dots, a_n) \mapsto a_i$  are the respective  $i^{\text{th}}$  projections.

Clearly,  $\tau_i\varphi_i = \eta\pi_i$  for every  $i \in [n]$ . Since  $\eta$  is surjective, we may define a mapping  $\psi_0 : X \rightarrow T(\Sigma_1, X_1) \times \dots \times T(\Sigma_n, X_n)$  such that  $x\psi_0\eta = x\varphi$  for every

$x \in X$ . If  $\psi : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{K}(\mathcal{T}_1, \dots, \mathcal{T}_n)$  is the homomorphic extension of  $\psi_0$ , then  $\psi\eta = \varphi$ .

Now,  $T$  is the union of finitely many sets  $\mathbf{a}\varphi^{-1}$  with  $\mathbf{a} = (a_1, \dots, a_n) \in F$ . Since  $\varphi\pi_i = \psi\tau_i\varphi_i$  for each  $i \in [n]$ , we have

$$\mathbf{a}\varphi^{-1} = \bigcap \{a_i(\varphi\pi_i)^{-1} \mid i \in [n]\} = \bigcap \{(a_i\varphi_i^{-1})(\psi\tau_i)^{-1} \mid i \in [n]\},$$

where each  $a_i\varphi_i^{-1}$  is a  $\theta_{U_i}$ -class, and therefore belongs to  $\mathcal{V}(\Sigma_i, X_i)$  by Proposition 8.2. By the definition of VUTs, this means that  $(a_i\varphi_i^{-1})(\psi\tau_i)^{-1} \in \mathcal{V}(\Sigma, X)$  for every  $i \in [n]$ , and hence also  $T \in \mathcal{V}(\Sigma, X)$ . This concludes the proof of  $\mathcal{V}^{at} \subseteq \mathcal{V}$ .  $\square$

The above results can be summed up as the following variety theorem.

**Theorem 9.5.** *The mappings  $\mathbf{VRA} \rightarrow \mathbf{VUT}$ ,  $\mathbf{K} \mapsto \mathbf{K}^t$ , and  $\mathbf{VUT} \rightarrow \mathbf{VRA}$ ,  $\mathcal{V} \mapsto \mathcal{V}^a$ , are mutually inverse isomorphisms between the lattices  $(\mathbf{VRA}, \subseteq)$  and  $(\mathbf{VUT}, \subseteq)$ .*  $\square$

## 10 Examples of varieties of unranked tree languages

We shall now introduce several varieties of unranked tree languages. Most of them correspond to some general variety of ranked tree languages considered in [25]. The following simple observations concerning g-morphisms of term algebras are helpful in many of the examples.

**Lemma 10.1.** *Let  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{T}_\Omega(Y)$  be a g-morphism.*

- (a)  $\text{hg}(f\varphi) = 0$  and  $\text{root}(f\varphi) = f\varphi = \iota(f)$  for every  $f \in \Sigma$ .
- (b) If  $t = f(t_1, \dots, t_m)$  ( $m > 0$ ), then  $t\varphi = \iota(f)(t_1\varphi, \dots, t_m\varphi)$  and  $\text{root}(t\varphi) = \iota(f)$ .
- (c)  $\text{hg}(t\varphi) \geq \text{hg}(t)$  for every  $t \in T_\Sigma(X)$ .
- (d)  $t\varphi^{-1}$  is finite for every  $t \in T_\Omega(Y)$ .  $\square$

All statements of Lemma 10.1 are obvious. Note, however, that (d) does not follow directly from (c), as in the ranked case.

### 10.1 Nilpotent unranked tree languages

For any  $\Sigma$  and  $X$ , let  $\text{Nil}(\Sigma, X)$  consist of all finite  $\Sigma X$ -tree languages and their complements in  $T_\Sigma(X)$ , and let  $\text{Nil} := \{\text{Nil}(\Sigma, X)\}$ . In view of Proposition 7.5(a), to prove that  $\text{Nil} \subseteq \text{Rec}$ , it suffices to note that, for all  $\Sigma$  and  $X$ , each singleton set  $\{t\} \subseteq T_\Sigma(X)$  is recognizable. It is clear that each set  $\text{Nil}(\Sigma, X)$  is closed under all Boolean operations, and condition (V4) and (V5) follow from

the facts that the pre-images  $p^{-1}(T)$  and  $T\varphi^{-1}$  are finite for any finite  $T$ ; for  $p^{-1}(T)$  this is obvious and for  $T\varphi^{-1}$  it follows from Lemma 10.1(d).

Similarly as in the unranked case [24, 26], it is easy to find the VRA corresponding to  $Nil$ . However, nilpotent unranked algebras cannot be defined just in terms of the height of trees as there are infinitely many trees of any height  $\geq 1$ . Let us define the *size*  $size(t)$  of tree  $t \in T_\Sigma(X)$  as the number of nodes of  $t$ , i.e.,

- (1)  $size(t) = 1$  for  $t \in \Sigma \cup X$ , and
- (2)  $size(t) = size(t_1) + \dots + size(t_m) + 1$  for  $t = f(t_1, \dots, t_m)$ .

We call an unranked algebra  $\mathcal{A} = (A, \Sigma)$  *nilpotent* if there exist an element  $a_0 \in A$  and a  $k \geq 1$  such that for any  $X$  and  $t \in T_\Sigma(X)$ , if  $size(t) \geq k$ , then  $t^{\mathcal{A}}(\alpha) = a_0$  for every  $\alpha : X \rightarrow A$ . The element  $a_0$  is then called the *absorbing state* of  $\mathcal{A}$  and the least  $k$  for which the above condition holds is its *degree (of nilpotency)*. For each  $k \geq 1$ , let  $\mathbf{Nil}_k$  denote the class of regular nilpotent algebras of degree  $\leq k$ , and let  $\mathbf{Nil}$  be the class of all regular nilpotent algebras. Obviously,  $\mathbf{Nil}_1 \subset \mathbf{Nil}_2 \subset \mathbf{Nil}_3 \subset \dots$  and  $\mathbf{Nil} = \bigcup_{k \geq 1} \mathbf{Nil}_k$ . Let us now show that each  $\mathbf{Nil}_k$ , and hence also  $\mathbf{Nil}$ , is a VRA. For this consider any  $k \geq 1$  and any regular algebras  $\mathcal{A} = (A, \Sigma)$  and  $\mathcal{B} = (B, \Omega)$ , and assume that  $\mathcal{A} \in \mathbf{Nil}_k$  with absorbing state  $a_0$ .

It is obvious that if  $\mathcal{B}$  is a g-subalgebra of  $\mathcal{A}$ , then also  $\mathcal{B} \in \mathbf{Nil}_k$  with  $a_0$  as the absorbing state. Next, let  $(\iota, \varphi) : \mathcal{A} \rightarrow \mathcal{B}$  be a g-epimorphism. For any  $X$  and any  $\beta : X \rightarrow B$ , there is a mapping  $\alpha : X \rightarrow A$  such that  $\alpha\varphi = \beta$ . Recall the mapping  $\iota_X : T_\Sigma(X) \rightarrow T_\Omega(X)$  defined in Section 3. It is easy to verify by induction on  $s$  that  $s^{\mathcal{A}}(\alpha)\varphi = \iota_X(s)^{\mathcal{B}}(\beta)$  for every  $s \in T_\Sigma(X)$ . Now, let  $t$  be any  $\Omega X$ -tree of size  $\geq k$ . As  $\iota$  is surjective, there exists an  $s \in T_\Sigma(X)$  such that  $\iota_X(s) = t$ , and hence  $t^{\mathcal{B}}(\beta) = \iota_X(s)^{\mathcal{B}}(\beta) = s^{\mathcal{A}}(\alpha)\varphi = a_0\varphi$ . This shows that  $\mathcal{B}$  is in  $\mathbf{Nil}_k$  with  $a_0\varphi$  as its absorbing state.

Assume now that also  $\mathcal{B} \in \mathbf{Nil}_k$  and let  $b_0$  be the absorbing state. Consider any g-product  $\varkappa(\mathcal{A}, \mathcal{B}) = (A \times B, \Gamma)$ , and any  $X$  and  $\gamma : X \rightarrow A \times B$ . Let us define  $\alpha : X \rightarrow A$ ,  $\beta : X \rightarrow B$ ,  $\iota : \Gamma \rightarrow \Sigma$  and  $\lambda : \Gamma \rightarrow \Omega$  by  $\alpha := \gamma\pi_1$ ,  $\beta := \gamma\pi_2$ ,  $\iota := \varkappa\pi_1$  and  $\lambda := \varkappa\pi_2$ , respectively. If  $t \in T_\Gamma(X)$ , then  $\iota_X(t) \in T_\Sigma(X)$ ,  $\lambda_X(t) \in T_\Omega(X)$  and  $size(\iota_X(t)) = size(\lambda_X(t)) = size(t)$ , and it can be verified by induction on  $t$  that  $t^{\varkappa(\mathcal{A}, \mathcal{B})}(\gamma) = (\iota_X(t)^{\mathcal{A}}(\alpha), \lambda_X(t)^{\mathcal{B}}(\beta))$ . This means that if  $size(t) \geq k$ , then  $t^{\varkappa(\mathcal{A}, \mathcal{B})}(\gamma) = (a_0, b_0)$  which shows that  $\varkappa(\mathcal{A}, \mathcal{B}) \in \mathbf{Nil}_k$ .

It remains to be shown that  $\mathbf{Nil}^t = Nil$ . First, let  $T \subseteq T_\Sigma(X)$  be recognized by some  $\mathcal{A} = (A, \Sigma)$  in  $\mathbf{Nil}_k$ , i.e.,  $T = F\varphi^{-1}$  for some morphism  $\varphi : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{A}$  and  $F \subseteq A$ . If  $a_0$  is the absorbing state of  $\mathcal{A}$  and  $\alpha : X \rightarrow A$  is the restriction of  $\varphi$  to  $X$ , then  $t^{\mathcal{A}}(\alpha) = a_0$  for every  $t \in T_\Sigma(X)$  of size  $\geq k$ . This means that  $T$  is finite if  $a_0 \notin F$  and co-finite if  $a_0 \in F$ . Hence,  $\mathbf{Nil}^t \subseteq Nil$ .

To prove the converse inclusion, consider any  $\Sigma$ ,  $X$  and a finite  $\Sigma X$ -tree language  $T$ . Let  $k := \max\{size(t) \mid t \in T\} + 1$  (for  $T = \emptyset$ , let  $k = 1$ ). We construct a nilpotent algebra  $\mathcal{A} = (A, \Sigma)$  recognizing  $T$  as follows. Let  $B := \{t \in T_\Sigma(X) \mid size(t) < k\}$  and  $A := B \cup \{a_0\}$  (with  $a_0 \notin B$ ), and for all  $f \in \Sigma$ ,

$m \geq 0$  and  $b_1, \dots, b_m \in A$  set

$$f_{\mathcal{A}}(b_1, \dots, b_m) = \begin{cases} f(b_1, \dots, b_m) & \text{if } f(b_1, \dots, b_m) \in B; \\ a_0 & \text{otherwise.} \end{cases}$$

It is clear that  $t^{\mathcal{A}}(\alpha) = a_0$  for every  $\alpha : X \rightarrow A$  whenever  $t \in T_{\Sigma}(X)$  and  $\text{size}(t) \geq k$ , and also that  $\mathcal{A}$  is regular. If  $\varphi : \mathcal{T}_{\Sigma}(X) \rightarrow \mathcal{A}$  is the morphism such that  $x\varphi = x$  for every  $x \in X$ , then  $t\varphi = t^{\mathcal{A}}(\alpha) = t$  if  $\text{size}(t) < k$  and  $t\varphi = a_0$  otherwise. This means that  $T = T\varphi^{-1}$ . For a co-finite  $T$ , we construct such an  $\mathcal{A}$  for  $S := T_{\Sigma}(X) \setminus T$  and obtain  $T$  as  $(A \setminus S)\varphi^{-1}$ . Hence,  $\text{Nil} \subseteq \mathbf{Nil}^t$ .

The above findings may be summed up as follows.

**Proposition 10.2.** *Nil is the VUT corresponding to the VRA Nil.*  $\square$

## 10.2 Definite unranked tree languages

The  $k$ -root  $\text{rt}_k(t)$  of a  $\Sigma X$ -tree  $t$  is defined as follows:

- (0)  $\text{rt}_0(t) = \varepsilon$ , where  $\varepsilon$  represents the empty root segment, for all  $t \in T_{\Sigma}(X)$ ;
- (1)  $\text{rt}_1(t) = \text{root}(t)$  for every  $t \in T_{\Sigma}(X)$ ;
- (2) for  $k \geq 2$ ,  $\text{rt}_k(t) = t$  if  $\text{hg}(t) < k$ , and  $\text{rt}_k(t) = f(\text{rt}_{k-1}(t_1), \dots, \text{rt}_{k-1}(t_m))$  if  $\text{hg}(t) \geq k$  and  $t = f(t_1, \dots, t_m)$ .

We call a recognizable unranked  $\Sigma X$ -tree language  $T$  *k-definite* if for all  $s, t \in T_{\Sigma}(X)$ , if  $\text{rt}_k(s) = \text{rt}_k(t)$  and  $s \in T$ , then  $t \in T$ , and it is *definite* if it is *k-definite* for some  $k \geq 0$ . Let  $\text{Def}_k = \{\text{Def}_k(\Sigma, X)\}$  and  $\text{Def} = \{\text{Def}(\Sigma, X)\}$  be the families of *k-definite* ( $k \geq 0$ ) and all definite tree languages, respectively. Clearly  $\text{Def}_0 \subset \text{Def}_1 \subset \text{Def}_2 \subset \dots$  and  $\text{Def} = \bigcup_{k \geq 0} \text{Def}_k$ . We could naturally verify directly that the families  $\text{Def}_k$  satisfy conditions (V1)–(V5), but let us show how they are obtained from consistent systems of congruences. For any  $k \geq 0$ ,  $\Sigma$  and  $X$ , define the relation  $\delta_k(\Sigma, X)$  in  $T_{\Sigma}(X)$  by

$$s \delta_k(\Sigma, X) t :\Leftrightarrow \text{rt}_k(s) = \text{rt}_k(t) \quad (s, t \in T_{\Sigma}(X)).$$

Note that for every  $k \geq 2$ , there are infinitely many  $\delta_k(\Sigma, X)$ -classes. Let  $\Delta(k) := \{\delta_k(\Sigma, X)\}_{\Sigma, X}$ . The following technical lemma is needed for showing that the families  $\text{Def}_k$  are VUTs.

**Lemma 10.3.** *Let  $(\iota, \varphi) : \mathcal{T}_{\Sigma}(X) \rightarrow \mathcal{T}_{\Omega}(Y)$  be a  $g$ -morphism. Then  $\text{rt}_k(t\varphi) = \text{rt}_k(\text{rt}_k(t)\varphi)$  for all  $t \in T_{\Sigma}(X)$  and  $k \geq 1$ .*

*Proof.* We proceed by induction on  $k \geq 1$ . The case  $k = 1$  is obvious:  $\text{rt}_1(t\varphi) = \text{root}(t\varphi) = \text{root}(\text{root}(t)\varphi) = \text{rt}_1(\text{rt}_1(t)\varphi)$ .

Assume now that  $k \geq 2$  and that the lemma holds for all smaller values of  $k$ . If  $\text{hg}(t) < k$ , then  $\text{rt}_k(t) = t$  and hence  $\text{rt}_k(\text{rt}_k(t)\varphi) = \text{rt}_k(t\varphi)$ . Assume that



$\text{hg}(t) \geq k$  and let  $t = f(t_1, \dots, t_m)$ . Then  $t\varphi = \iota(f)(t_1\varphi, \dots, t_m\varphi)$  and therefore

$$\begin{aligned} \text{rt}_k(t\varphi) &= \iota(f)(\text{rt}_{k-1}(t_1\varphi), \dots, \text{rt}_{k-1}(t_m\varphi)) \\ &= \iota(f)(\text{rt}_{k-1}(\text{rt}_{k-1}(t_1)\varphi), \dots, \text{rt}_{k-1}(\text{rt}_{k-1}(t_m)\varphi)) \\ &= \text{rt}_k(\iota(f)(\text{rt}_{k-1}(t_1)\varphi, \dots, \text{rt}_{k-1}(t_m)\varphi)) \\ &= \text{rt}_k(f(\text{rt}_{k-1}(t_1), \dots, \text{rt}_{k-1}(t_m)))\varphi \\ &= \text{rt}_k(\text{rt}_k(t)\varphi), \end{aligned}$$

where we also used the inductive assumption.  $\square$

**Proposition 10.4.** *For every  $k \geq 0$ ,  $\Delta(k)$  is a consistent system of congruences, and  $\text{Def}_k$  is the quasi-principal VUT defined by it, and hence also  $\text{Def}$  is a VUT.*

*Proof.* Obviously it suffices to show that the following statements (a)–(c) hold for every  $k \geq 0$  and for all alphabets  $\Sigma, \Omega, X, Y$ .

- (a)  $\delta_k(\Sigma, X)$  is a congruence of  $\mathcal{T}_\Sigma(X)$ .
- (b)  $\delta_k(\Sigma, X) \subseteq \varphi \circ \delta_k(\Omega, Y) \circ \varphi^{-1}$  for any g-morphism  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{T}_\Omega(Y)$ .
- (c) A recognizable  $\Sigma X$ -tree language  $T$  is  $k$ -definite iff  $\delta_k(\Sigma, X) \subseteq \theta_T$ .

Since  $\delta_0(\Sigma, X) = \nabla_{T_\Sigma(X)}$ , statement (a) trivially holds for  $k = 0$ . For  $k > 0$ , it follows from the obvious fact that  $\text{rt}_k(s) = \text{rt}_k(t)$  implies  $\text{rt}_{k-1}(s) = \text{rt}_{k-1}(t)$ .

To prove (b) it suffices to show that if  $s, t \in T_\Sigma(X)$  and  $\text{rt}_k(s) = \text{rt}_k(t)$ , then  $\text{rt}_k(s\varphi) = \text{rt}_k(t\varphi)$ , and this follows from Lemma 10.3.

Let  $T \in \text{Rec}(\Sigma, X)$ . Since  $T$  being  $k$ -definite means precisely that  $T$  is saturated by  $\delta_k(\Sigma, X)$ , statement (c) follows from Lemma 6.2.

As the union of the chain  $\text{Def}_0 \subset \text{Def}_1 \subset \text{Def}_2 \subset \dots$ , also  $\text{Def}$  is a VUT.  $\square$

### 10.3 Reverse definite unranked tree languages

A  $\Sigma X$ -tree  $s$  is a *subtree* of a  $\Sigma X$ -tree  $t$  if  $t = p(s)$  for some context  $p \in C_\Sigma(X)$ . For any  $t \in T_\Sigma(X)$ , let  $\text{st}(t)$  denote the set of subtrees of  $t$ , and for each  $k \geq 0$ , let  $\text{st}_k(t) = \{s \in \text{st}(t) \mid \text{hg}(s) < k\}$ . Note that  $\text{st}_0(t) = \emptyset$  for every  $t$ .

We call a recognizable unranked  $\Sigma X$ -tree language  $T$  *reverse  $k$ -definite* if for all  $s, t \in T_\Sigma(X)$ , if  $\text{st}_k(s) = \text{st}_k(t)$  and  $s \in T$ , then  $t \in T$ , and it is *reverse definite* if it is reverse  $k$ -definite for some  $k \geq 0$ . Let  $R\text{Def}_k = \{R\text{Def}_k(\Sigma, X)\}$  and  $R\text{Def} = \{R\text{Def}(\Sigma, X)\}$  be the families of  $k$ -reverse definite ( $k \geq 0$ ) and all reverse definite tree languages. Clearly  $R\text{Def}_0 \subset R\text{Def}_1 \subset R\text{Def}_2 \subset \dots$  and  $R\text{Def} = \bigcup_{k \geq 0} R\text{Def}_k$ .

For each  $k \geq 0$ , a consistent system of congruences  $P(k) = \{\rho_k(\Sigma, X)\}_{\Sigma, X}$  defining  $R\text{Def}_k$  is obtained when we set for any  $\Sigma$  and  $X$ ,

$$s \rho_k(\Sigma, X) t :\Leftrightarrow \text{st}_k(s) = \text{st}_k(t) \quad (s, t \in T_\Sigma(X)).$$

To prove the consistency of the systems  $P(k)$ , we need the following fact.

**Lemma 10.5.** *Let  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{T}_\Omega(Y)$  be a  $g$ -morphism. Then  $\text{st}_k(t\varphi) = \bigcup \{\text{st}_k(s\varphi) \mid s \in \text{st}_k(t)\}$  for all  $t \in T_\Sigma(X)$  and  $k \geq 0$ .*

*Proof.* The equality is obvious when  $k = 0$ , so we assume that  $k > 0$  and proceed by induction on  $t \in T_\Sigma(X)$ .

If  $t \in \Sigma \cup X$ , then the equality certainly holds because  $\text{st}_k(t) = \{t\}$ .

Let  $t = f(t_1, \dots, t_m)$ , where  $m > 0$ , and assume that the claim holds for all trees of height  $< \text{hg}(t)$ . If  $\text{hg}(t) < k$ , then again  $t \in \text{st}_k(t)$ . If  $\text{hg}(t) \geq k$ , then  $\text{st}_k(t) = \text{st}_k(t_1) \cup \dots \cup \text{st}_k(t_m)$ . Since we also have  $t\varphi = \iota(f)(t_1\varphi, \dots, t_m\varphi)$ , we get

$$\begin{aligned} \text{st}_k(t\varphi) &= \text{st}_k(t_1\varphi) \cup \dots \cup \text{st}_k(t_m\varphi) \\ &= \bigcup \{\text{st}_k(s\varphi) \mid s \in \text{st}_k(t_1)\} \cup \dots \cup \bigcup \{\text{st}_k(s\varphi) \mid s \in \text{st}_k(t_m)\} \\ &= \bigcup \{\text{st}_k(s\varphi) \mid s \in \text{st}_k(t)\} \end{aligned}$$

by applying the inductive assumption to the trees  $t_1, \dots, t_m$ .  $\square$

**Proposition 10.6.** *For every  $k \geq 0$ ,  $P(k)$  is a consistent system of congruences, and  $RDef_k$  is the quasi-principal VUT defined by it, and hence also  $RDef$  is a VUT.*

*Proof.* We should show that the following hold for all  $k \geq 0$ ,  $\Sigma$ ,  $\Omega$ ,  $X$ , and  $Y$ .

- (a)  $\rho_k(\Sigma, X)$  is a congruence of  $\mathcal{T}_\Sigma(X)$ .
- (b)  $\rho_k(\Sigma, X) \subseteq \varphi \circ \rho_k(\Omega, Y) \circ \varphi^{-1}$  for any  $g$ -morphism  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{T}_\Omega(Y)$ .
- (c) A recognizable  $\Sigma X$ -tree language  $T$  is reverse  $k$ -definite iff  $\rho_k(\Sigma, X) \subseteq \theta_T$ .

Statement (a) holds for  $k = 0$  as  $\rho_0(\Sigma, X) = \nabla_{T_\Sigma(X)}$ . Let  $k > 0$  and consider any  $f \in \Sigma$ ,  $m > 0$  and any  $s_1, \dots, s_m, t_1, \dots, t_m \in T_\Sigma(X)$  such that  $(s_1, t_1), \dots, (s_m, t_m) \in \rho_k(\Sigma, X)$ . We distinguish two cases. If  $\text{hg}(f(s_1, \dots, s_m)) < k$ , then  $\text{hg}(s_1), \dots, \text{hg}(s_m) < k$  and we must have  $s_1 = t_1, \dots, s_m = t_m$ , from which  $f(s_1, \dots, s_m) \rho_k(\Sigma, X) f(t_1, \dots, t_m)$  trivially follows. On the other hand, if  $\text{hg}(f(s_1, \dots, s_m)) \geq k$ , then  $\text{st}_k(s_1) = \text{st}_k(t_1), \dots, \text{st}_k(s_m) = \text{st}_k(t_m)$  implies

$$\text{st}_k(f(s_1, \dots, s_m)) = \text{st}_k(s_1) \cup \dots \cup \text{st}_k(s_m) = \text{st}_k(f(t_1, \dots, t_m)),$$

and hence again  $f(s_1, \dots, s_m) \rho_k(\Sigma, X) f(t_1, \dots, t_m)$ .

To prove (b) it suffices to show that if  $\text{st}_k(s) = \text{st}_k(t)$  for some  $s, t \in T_\Sigma(X)$ , then  $\text{st}_k(s\varphi) = \text{st}_k(t\varphi)$ , but this holds by Lemma 10.5.

Let  $T \in \text{Rec}(\Sigma, X)$ . Since  $T$  being reverse  $k$ -definite means precisely that  $T$  is saturated by  $\rho_k(\Sigma, X)$ , statement (c) follows from Lemma 6.2.

As the union of the chain  $RDef_0 \subset RDef_1 \subset RDef_2 \subset \dots$ , also  $RDef$  is a VUT.  $\square$

## 10.4 Generalized definite tree languages

For any  $h, k \geq 0$ , we call an unranked  $\Sigma X$ -tree language  $T$   *$h, k$ -definite* if for all  $s, t \in T_\Sigma(X)$ , if  $\text{st}_h(s) = \text{st}_h(t)$  and  $\text{rt}_k(s) = \text{rt}_k(t)$ , then  $s \in T$  iff  $t \in T$ , and it is *generalized definite* if it is  $h, k$ -definite for some  $h, k \geq 0$ . Let  $GDef_{h,k} = \{GDef_{h,k}(\Sigma, X)\}$  and  $GDef = \{GDef(\Sigma, X)\}$  be the families of all recognizable  $h, k$ -definite ( $h, k \geq 0$ ) and all recognizable general definite tree languages. Clearly  $GDef_{h,k} \subseteq GDef_{h',k'}$  whenever  $h \leq h'$  and  $k \leq k'$ , and  $GDef = \bigcup_{h,k \geq 0} GDef_{h,k}$ .

For any  $h, k \geq 0$ ,  $\Sigma$  and  $X$ , let  $\gamma_{h,k}(\Sigma, X) = \rho_h(\Sigma, X) \cap \delta_k(\Sigma, X)$ , and let  $\Gamma(h, k) := \{\gamma_{h,k}(\Sigma, X)\}_{\Sigma, X}$ . The following proposition can be proved simply by combining the arguments used in the previous two examples.

**Proposition 10.7.** *For all  $h, k \geq 0$ ,  $\Gamma(h, k)$  is a consistent system of congruences, and  $GDef_{h,k}$  is the quasi-principal VUT defined by it, and hence  $GDef$  is also a VUT.  $\square$*

## 10.5 Locally testable unranked tree languages

For any  $k \geq 2$ ,  $\Sigma$  and  $X$ , we define the set  $\text{fork}_k(t)$  of  *$k$ -forks* of a  $\Sigma X$ -tree  $t$  thus:

- (1) if  $\text{hg}(t) < k - 1$ , then  $\text{fork}_k(t) = \emptyset$ ;
- (2) if  $\text{hg}(t) \geq k - 1$  and  $t = f(t_1, \dots, t_m)$ , then  $\text{fork}_k(t) = \{\text{rt}_k(t)\} \cup \text{fork}_k(t_1) \cup \dots \cup \text{fork}_k(t_m)$ .

Clearly,  $\text{fork}_k(t)$  is a finite set of  $\Sigma X$ -trees of height  $k - 1$ . For example, if  $t = f(x, f(y))$ , then  $\text{fork}_2(t) = \{f(x, f), f(y)\}$ ,  $\text{fork}_3(t) = \{t\}$  and  $\text{fork}_k(t) = \emptyset$  for all  $k \geq 4$ . Note that the set of all possible  $k$ -forks of  $\Sigma X$ -trees is infinite.

Now, let  $\lambda_k(\Sigma, X)$  be the relation on  $T_\Sigma(X)$  such that for any  $s, t \in T_\Sigma(X)$ ,

$$s \lambda_k(\Sigma, X) t :\Leftrightarrow \text{st}_{k-1}(s) = \text{st}_{k-1}(t), \text{rt}_{k-1}(s) = \text{rt}_{k-1}(t), \text{fork}_k(s) = \text{fork}_k(t).$$

It is easy to see that  $\lambda_k(\Sigma, X) \in \text{Con}(\mathcal{T}_\Sigma(X))$ . An unranked  $\Sigma X$ -tree language is said to be  *$k$ -testable* if it is saturated by  $\lambda_k(\Sigma, X)$ , and it is called *locally testable* if it is  $k$ -testable for some  $k \geq 2$ . Let  $Loc_k(\Sigma, X)$  be the set of all recognizable  $k$ -testable  $\Sigma X$ -tree languages, and let  $Loc(\Sigma, X) := \bigcup_{k \geq 2} Loc_k(\Sigma, X)$  be the set of all recognizable locally testable  $\Sigma X$ -tree languages.

Note that for any  $\Sigma X$ -tree  $t$  of height  $\geq k - 1$ ,  $\text{st}_{k-1}(t)$  consists of the subtrees of  $t$  of height  $\leq k - 2$ ,  $\text{rt}_{k-1}(t)$  is its root segment of height  $k - 2$ , and  $\text{fork}_k(t)$  consists of its forks of height  $k - 1$ . In particular, if  $t$  is a string represented as a unary tree, then  $\text{st}_{k-1}(t)$  consists of the prefixes of  $t$  of length  $\leq k - 1$ ,  $\text{rt}_{k-1}(t)$  is the suffix of  $t$  of length  $k - 1$ , and  $\text{fork}_k(t)$  is the set of its substrings of length  $k$ . Hence, our unranked  $k$ -testable tree languages are obtained by a natural adaptation of the usual definition of  $k$ -testable string languages (cf. [11], for example).

To show that the families  $Loc_k := \{Loc_k(\Sigma, X)\}$  ( $k \geq 2$ ) and  $Loc := \{Loc(\Sigma, X)\}$  are varieties, we consider the systems of congruences  $\Lambda(k) :=$

$\{\lambda_k(\Sigma, X)\}_{\Sigma, X}$  ( $k \geq 2$ ). For proving the consistency of these systems, we need the following lemma.

**Lemma 10.8.** *If  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{T}_\Omega(Y)$  is a g-morphism and  $k \geq 2$ , then*

$$\text{fork}_k(t\varphi) = \bigcup \{\text{rt}_k(u\varphi) \mid u \in \text{fork}_k(t)\} \cup \bigcup \{\text{fork}_k(s\varphi) \mid s \in \text{st}_{k-1}(t)\},$$

for every  $t \in T_\Sigma(X)$ .

*Proof.* In the course of this proof, we use a couple of times the obvious fact that if  $s$  is a subtree of  $s'$ , then  $\text{fork}_k(s) \subseteq \text{fork}_k(s')$  for all  $k \geq 2$ . Let *RHS* denote the righthand side of the claimed equality. We proceed by induction on the height of  $t \in T_\Sigma(X)$ .

If  $\text{hg}(t) < k-1$ , then  $\text{fork}_k(t) = \emptyset$ . Moreover,  $\text{fork}_k(s\varphi) \subseteq \text{fork}_k(t\varphi)$  for every  $s \in \text{st}_{k-1}(t)$  as  $s \in \text{st}(t)$  clearly implies  $s\varphi \in \text{st}(t\varphi)$ . Hence, *RHS*  $\subseteq \text{fork}_k(t\varphi)$ . On the other hand,  $\text{fork}_k(t\varphi) \subseteq \text{RHS}$  because now  $t \in \text{st}_{k-1}(t)$ .

Let  $\text{hg}(t) \geq k-1$  and let  $t = f(t_1, \dots, t_m)$ , and assume that the equality holds for all trees of lesser height. Then  $t\varphi = \iota(f)(t_1\varphi, \dots, t_m\varphi)$ .

To prove the inclusion  $\text{fork}_k(t\varphi) \subseteq \text{RHS}$ , consider any  $v \in \text{fork}_k(t\varphi)$ . Since  $\text{fork}_k(t\varphi) = \{\text{rt}_k(t\varphi)\} \cup \text{fork}_k(t_1\varphi) \cup \dots \cup \text{fork}_k(t_m\varphi)$ , there are two possibilities. If  $v = \text{rt}_k(t\varphi)$ , then  $v = \text{rt}_k(\text{rt}_k(t)\varphi)$  by Lemma 10.3, and hence  $v \in \text{RHS}$  as  $\text{rt}_k(t) \in \text{fork}_k(t)$ . If  $v \in \text{fork}_k(t_i\varphi)$  for some  $i \in [m]$ , then by the induction assumption, either  $v = \text{rt}_k(u\varphi)$  for some  $u \in \text{fork}_k(t_i) (\subseteq \text{fork}_k(t))$  or  $v \in \text{fork}_k(s\varphi)$  for some  $s \in \text{st}_{k-1}(t_i) (\subseteq \text{st}_{k-1}(t))$ . In either case,  $v \in \text{RHS}$ .

Assume now that  $v \in \text{RHS}$ . If  $v = \text{rt}_k(u\varphi)$  for some  $u \in \text{fork}_k(t)$ , we have two cases to consider:  $u = \text{rt}_k(t)$  or  $u \in \text{fork}_k(t_i)$  for some  $i \in [m]$ . In the first case,  $v = \text{rt}_k(\text{rt}_k(t)\varphi) = \text{rt}_k(t\varphi) \in \text{fork}_k(t\varphi)$  by Lemma 10.3. In the second case,  $v \in \text{fork}_k(t_i\varphi) (\subseteq \text{fork}_k(t\varphi))$  by the inductive hypothesis.

Finally, if  $v \in \text{fork}_k(s\varphi)$  for some  $s \in \text{st}_{k-1}(t)$ , then  $s \in \text{st}_{k-1}(t_i)$  for some  $i \in [m]$ , and hence  $v \in \text{fork}_k(t_i\varphi) \subseteq \text{fork}_k(t\varphi)$  by the induction assumption. This completes the proof of the inclusion *RHS*  $\subseteq \text{fork}_k(t\varphi)$ .  $\square$

**Proposition 10.9.** *For every  $k \geq 2$ , the system of congruences  $\Lambda(k)$  is consistent, and  $\text{Loc}_k$  is the quasi-principal VUT defined by it, and hence also  $\text{Loc}$  is a VUT.*

*Proof.* Consider any  $k \geq 2$  and any g-morphism  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{T}_\Omega(Y)$ , and let  $s, t \in T_\Sigma(X)$  be such that  $s \lambda_k(\Sigma, X) t$ . To prove the consistency of  $\Lambda(k)$ , we should show that  $s\varphi \lambda_k(\Omega, Y) t\varphi$ .

By the proofs of Propositions 10.6 and 10.4, we know that  $\text{st}_{k-1}(s\varphi) = \text{st}_{k-1}(t\varphi)$  and  $\text{rt}_{k-1}(s\varphi) = \text{rt}_{k-1}(t\varphi)$  follow from  $\text{st}_{k-1}(s) = \text{st}_{k-1}(t)$  and  $\text{rt}_{k-1}(s) = \text{rt}_{k-1}(t)$ , respectively. Similarly,  $\text{fork}_k(s) = \text{fork}_k(t)$  and  $\text{st}_{k-1}(s) = \text{st}_{k-1}(t)$  imply  $\text{fork}_k(s\varphi) = \text{fork}_k(t\varphi)$  by Lemma 10.8. Hence  $s\varphi \lambda_k(\Omega, Y) t\varphi$ .

That  $\text{Loc}_k$  is the quasi-principal VUT defined by  $\Lambda(k)$  follows immediately from its definition. Finally,  $\text{Loc}$  is a VUT as the union of the chain  $\text{Loc}_2 \subseteq \text{Loc}_3 \subseteq \dots$   $\square$

## 10.6 Aperiodic tree languages

To show that the natural unranked counterparts of the aperiodic tree languages [31] form a variety is as easy as in the ranked case [25].

For any  $p, q \in C_\Sigma(X)$  and  $t \in T_\Sigma(X)$ , let  $p \cdot q := q(p)$  and  $t \cdot p := p(t)$ . Obviously,  $(C_\Sigma(X), \cdot, \xi)$  is a monoid and the powers  $p^n$  ( $n \geq 0$ ) of a  $\Sigma X$ -context  $p$  are defined as usual. An unranked tree language  $T \subseteq T_\Sigma(X)$  is called *aperiodic* (or *noncounting*) if there exists a number  $n \geq 0$  such that for all  $q, r \in C_\Sigma(X)$  and  $t \in T_\Sigma(X)$ ,

$$t \cdot q^{n+1} \cdot r \in T \Leftrightarrow t \cdot q^n \cdot r \in T.$$

If  $T$  is aperiodic, the least  $n$  for which the above condition holds, is denoted by  $\text{ia}(T)$ . Let  $\text{Ap}(\Sigma, X)$  be the set of all recognizable aperiodic  $\Sigma X$ -tree languages. Let  $\text{Ap} := \{\text{Ap}(\Sigma, X)\}$ .

**Proposition 10.10.** *Ap is a VUT.*

*Proof.* It is obvious that  $\text{Ap}$  satisfies conditions (V1)–(V3). To verify (V4), consider any  $T \in \text{Ap}(\Sigma, X)$  and  $p \in C_\Sigma(X)$ . If  $\text{ia}(T) = n$ , then for all  $q, r \in C_\Sigma(X)$  and  $t \in T_\Sigma(X)$ ,

$$t \cdot q^{n+1} \cdot r \in p^{-1}(T) \Leftrightarrow t \cdot q^{n+1} \cdot (r \cdot p) \in T \Leftrightarrow t \cdot q^n \cdot (r \cdot p) \in T \Leftrightarrow t \cdot q^n \cdot r \in p^{-1}(T),$$

which shows that  $p^{-1}(T) \in \text{Ap}(\Sigma, X)$ .

Finally, let  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{T}_\Omega(Y)$  be a g-morphism,  $T \in \text{Ap}(\Omega, Y)$  and  $\text{ia}(T) = n$ . Define  $\hat{\varphi} : C_\Sigma(X) \rightarrow C_\Omega(Y)$  as follows:

- (1)  $\xi \hat{\varphi} := \xi$ ;
- (2) if  $p = f(t_1, \dots, q, \dots, t_m)$  for some  $f \in \Sigma$ ,  $m \geq 1$ ,  $t_1, \dots, t_m \in T_\Sigma(X)$  and  $q \in C_\Sigma(X)$ , then  $p \hat{\varphi} := \iota(f)(t_1 \varphi, \dots, q \hat{\varphi}, \dots, t_m \varphi)$ .

It is easy to see that  $\hat{\varphi}$  is a monoid morphism and that  $(t \cdot p) \varphi = t \varphi \cdot p \hat{\varphi}$  for all  $t \in T_\Sigma(X)$  and  $p \in C_\Sigma(X)$ . This implies that, for all  $q, r \in C_\Sigma(X)$  and  $t \in T_\Sigma(X)$ ,

$$t \cdot q^{n+1} \cdot r \in T \varphi^{-1} \Leftrightarrow t \varphi \cdot (q \hat{\varphi})^{n+1} \cdot r \hat{\varphi} \in T \Leftrightarrow t \varphi \cdot (q \hat{\varphi})^n \cdot r \hat{\varphi} \in T \Leftrightarrow t \cdot q^n \cdot r \in T \varphi^{-1},$$

which shows that  $\text{Ap}$  satisfies (V5), too.  $\square$

## 10.7 Piecewise testable tree languages

As our final example, we consider piecewise testable unranked tree languages. As shown in [19], a natural definition of piecewise testable subtrees can be based on the well-known homeomorphic embedding order of trees (cf. [2], for example), and a corresponding order implicitly underlies the definition of the piecewise testable ‘forests’ (i.e., finite sequences of unranked trees) considered in [4]. The following presentation parallels that of [19] with the small modifications introduced in [27].

For any  $\Sigma$ ,  $X$  and  $k \geq 0$ , the *homeomorphic embedding order*  $\preceq$  on  $T_\Sigma(X)$  is defined by stipulating that for any  $s, t \in T_\Sigma(X)$ ,  $s \preceq t$  if and only if

- (1)  $s = t$ , or
- (2)  $s = f(s_1, \dots, s_m)$  and  $t = f(t_1, \dots, t_m)$  where  $s_1 \trianglelefteq t_1, \dots, s_m \trianglelefteq t_m$ , or
- (3)  $t = f(t_1, \dots, t_m)$  and  $s \trianglelefteq t_i$  for some  $i \in [m]$ .

Consider any  $\Sigma$ ,  $X$  and  $k \geq 0$ . For any  $t \in T_\Sigma(X)$ , let  $P_k(t) := \{s \in T_\Sigma(X) \mid s \trianglelefteq t, \text{hg}(s) < k\}$ , and then define  $\tau_k(\Sigma, X) := \{(s, t) \mid s, t \in T_\Sigma(X), P_k(s) = P_k(t)\}$ . Now, an unranked  $\Sigma X$ -tree language is said to be *piecewise  $k$ -testable* if it is saturated by  $\tau_k(\Sigma, X)$ , and it is *piecewise testable* if it is piecewise  $k$ -testable for some  $k \geq 0$ . Let  $Pwt_k(\Sigma, X)$  denote the set of all recognizable piecewise  $k$ -testable unranked  $\Sigma X$ -tree languages, and let  $Pwt(\Sigma, X) := \bigcup_{k \geq 0} Pwt_k(\Sigma, X)$  be the set of all recognizable piecewise testable unranked  $\Sigma X$ -tree languages. We want to prove that the families  $Pwt_k := \{Pwt_k(\Sigma, X)\}$  ( $k \geq 0$ ) and  $Pwt := \{Pwt(\Sigma, X)\}$  are VUTs. For this it suffices to show that for every  $k \geq 0$ ,  $T(k) := \{\tau_k(\Sigma, X)\}_{\Sigma, X}$  is a consistent system of congruences that defines  $Pwt_k$ .

It is easy to see that  $\tau_k(\Sigma, X) \in \text{Con}(\mathcal{T}_\Sigma(X))$  for every  $k \geq 0$ . (Note, however, that  $\tau_k(\Sigma, X)$  is not of finite index when  $k \geq 2$ .) For showing that the system  $T(k)$  is consistent, we need the following lemma.

**Lemma 10.11.** *Let  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{T}_\Omega(Y)$  be a  $g$ -morphism. For any  $k \geq 0$ ,  $s \in T_\Sigma(X)$  and  $t \in P_k(s\varphi)$ , there exists an  $s' \in P_k(s)$  such that  $t \in P_k(s'\varphi)$ .*

*Proof.* The proof goes by induction on  $k$ . The case  $k = 0$  is trivial since  $P_0(s\varphi)$  is empty. If  $t \in P_1(s\varphi)$ , then  $t \in \Omega \cup Y$ , and now we proceed by induction on  $s$ . If  $s \in \Sigma \cup X$ , we may let  $s'$  be  $s$ . If  $s = f(s_1, \dots, s_m)$ , then  $s\varphi = \iota(f)(s_1\varphi, \dots, s_m\varphi)$  and we must have  $t \in P_1(s_i\varphi)$  for some  $i \in [m]$ . By our tacit inductive assumption, there exists an  $s' \in P_1(s_i) \subseteq P_1(s)$  such that  $t \in P_1(s'\varphi)$ .

Assume now that  $k \geq 2$  and that the lemma holds for all smaller values of  $k$ . Again, we proceed by induction on  $s$ . If  $s \in \Sigma \cup X$ , then  $\text{hg}(s) < k$ , and we may set  $s' := s$ . Let  $s = f(s_1, \dots, s_m)$  and suppose the claim holds for all smaller trees. Since  $s\varphi = \iota(f)(s_1\varphi, \dots, s_m\varphi)$ , there are two cases to consider. If  $t \in P_k(s_i\varphi)$  for some  $i \in [m]$ , the required  $s'$  can be found as a piecewise subtree of  $s_i$ . Otherwise,  $t = \iota(f)(t_1, \dots, t_m)$  for some  $t_1 \in P_{k-1}(s_1\varphi), \dots, P_{k-1}(s_m\varphi)$ . By the main inductive assumption, there are trees  $s'_1 \in P_{k-1}(s_1), \dots, s'_m \in P_{k-1}(s_m)$  such that  $t_1 \in P_{k-1}(s'_1\varphi), \dots, t_m \in P_{k-1}(s'_m\varphi)$ . Then  $t \in P_k(s'\varphi)$  for  $s' := f(s'_1, \dots, s'_m) \in P_k(s)$ .  $\square$

**Proposition 10.12.** *For every  $k \geq 0$ , the system of congruences  $T(k)$  is consistent, and  $Pwt_k$  is the quasi-principal VUT defined by it, and hence also  $Pwt$  is a VUT.*

*Proof.* That the system  $T(k)$  is consistent follows directly from Lemma 10.11. Indeed, if  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{T}_\Omega(Y)$  is a  $g$ -morphism and  $s \tau_k(\Sigma, X) t$ , then  $s\varphi \tau_k(\Omega, Y) t\varphi$  because  $P_k(s) = P_k(t)$  implies  $P_k(s\varphi) = P_k(t\varphi)$  by that lemma.

By the definition of  $\tau_k(\Sigma, X)$ , a recognizable  $\Sigma X$ -tree language  $T$  is piecewise  $k$ -testable if and only if  $T$  is saturated by  $\tau_k(\Sigma, X)$ , and this is the case exactly in case  $\tau_k(\Sigma, X) \subseteq \theta_T$ . This means that  $\mathcal{V}_{T(k)} = Pwt_k$ .  $\square$

## 11 Concluding remarks

We have introduced and studied varieties of unranked tree languages that contain languages over all operator and leaf alphabets. We also defined the basic algebraic notions, such as subalgebras, morphisms and direct products, for unranked algebras in a way that allows us to consider algebras over any operator alphabets together. In particular, we have considered regular algebras, i. e., finite unranked algebras in which the operations are defined by regular languages. A bijective correspondence between varieties of unranked tree languages and varieties of regular algebras was established via syntactic algebras. We have also shown that the natural unranked counterparts of several known varieties of ranked tree languages form varieties in our sense. In many of these examples we made use of a general scheme by which so-called quasi-principal varieties are obtained from certain systems of congruences of term algebras.

Of course, much remains to be done. In particular, many of the example varieties considered here would deserve a deeper study. For example, it is natural to ask for characterizations of the corresponding varieties of regular algebras, or whether there are logics matching some of these varieties.

## References

- [1] J. Almeida, On pseudovarieties, varieties of languages, filters of congruences, pseudo-identities and related topics, *Algebra Universalis* 27 (1990) 333–350.
- [2] F. Baader, T. Nipkow, *Term Rewriting and All That*, Cambridge University Press, Cambridge, 1998.
- [3] C. Bergman, *Universal Algebra. Fundamentals and Selected Topics*, CRC Press, Taylor & Francis Group, Boca Raton, FL, 2012.
- [4] M. Bojańczyk, L. Segoufin, H. Straubing, Piecewise testable tree languages, *Log. Methods in Comput. Sci.* 8 (2012) 1–32.
- [5] M. Bojańczyk, I. Walukiewicz, Forest algebras, in: J. Flum, E. Grädel, T. Wilke (Eds.), *Automata and Logic: History and Perspectives*, Texts in Logic and Games, Vol. 2, Amsterdam University Press, Amsterdam 2008, pp. 107–132.
- [6] A. Brüggemann-Klein, M. Murata, D. Wood, Regular tree and hedge languages over unranked alphabets, Version 1, HKUST Theoretical Computer Science Center Research Report HKUST-TCS-2001-05, Hong Kong, 2001.
- [7] S. Burris, H.P. Sankappanavar, *A Course in Universal Algebra*, Springer-Verlag, New York, 1981.
- [8] H. Comon, M. Dauchet, R. Gilleron, C. Löding, F. Jacquemard, D. Lugiez, S. Tison, M. Tommasi, *Tree Automata Techniques and Applications*, available online: <http://www.grappa.univ-lille3.fr/tata>, 2007.

- [9] J. Cristau, C. Löding, W. Thomas, Deterministic automata on unranked trees, in: M. Liškiewicz, R. Reischuk (Eds.), *Fundamentals of Computation Theory*, Proc. 15th Internat. Symp. FCT 2005, Lübeck 2005, Lect. Notes Comput. Sci. 3623, Springer, Berlin, 2005, pp. 68–79.
- [10] B.A. Davey, H.A. Priestley, *Introduction to Lattices and Order*, second ed., Cambridge University Press, Cambridge 2002.
- [11] S. Eilenberg, *Automata, Languages, and Machines. Vol. B*, Academic Press, New York, 1976.
- [12] F. Gécseg, M. Steinby, *Tree Automata*, Akadémiai Kiadó, Budapest, 1984.
- [13] F. Gécseg, M. Steinby, Tree languages, in: G. Rozenberg, A. Salomaa (Eds.), *Handbook of Formal Languages*, Vol. 3, Springer, Berlin, 1997, pp. 1–69.
- [14] E. Graczyńska, D. Schweigert, Hypervarieties of a given type, *Algebra Universalis* 27 (1990) 303–318.
- [15] E. Jurvanen, A. Potthoff, W. Thomas, Tree languages recognizable by regular frontier check, in: G. Rozenberg, A. Salomaa (Eds.), *Developments in Language Theory*, World Scientific, Singapore, 1994, pp. 3–17.
- [16] W. Martens, J. Niehren, On the minimization of XML Schemas and tree automata for unranked trees, *J. Comput. Syst. Sci.* 73 (2007) 550–583.
- [17] F. Neven, Automata, Logic, and XML, in: J. Bradfield (Ed.), *Computer Science Logic*, Proc. 16th Internat. Workshop, CSL 2002, Edinburgh, UK, 2002, Lect. Notes Comput. Sci. 2471, Springer, Berlin, 2002, pp. 2–26.
- [18] C. Pair, A. Quere, Définition et étude de langages réguliers, *Inform. Control* 13(6) (1968) 565–593.
- [19] V. Piirainen, Piecewise testable tree languages, TUCS Technical Report 634, Turku Centre for Computer Science, Turku, 2004.
- [20] D. Schweigert, Hyperidentities, in: I.G. Rosenberg, G. Sabidussi (Eds.), *Algebras and Orders*, Kluwer Academic Publishers, Dordrecht, 1993, pp. 405–506.
- [21] T. Schwentick, Automata for XML – A survey, *J. Comput. Syst. Sci.* 73 (2007) 289–315.
- [22] L. Segoufin, V. Vianu, Validating streaming XML documents, in: L. Popa (Ed.), *ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, Proc. 21th Internat. Symp., PODS 2002, Wisconsin, USA, 2002, ACM, 2002, pp. 53–64.



- [23] M. Steinby, Syntactic algebras and varieties of recognizable sets, in: M.C. Gaudel, J.P. Jouannaud (Eds.), *Les Arbres en Algèbre et en Programmation*, Proc. 4th CAAP, Lille 1979, University of Lille, Lille, 1979, pp. 226–240.
- [24] M. Steinby, A theory of tree language varieties, in: M. Nivat, A. Podelski, (Eds.), *Tree Automata and Languages*, North-Holland, Amsterdam, 1992, pp. 57–81.
- [25] M. Steinby, General varieties of tree languages, *Theor. Comput. Sci.* 205 (1998) 1–43.
- [26] M. Steinby, Algebraic classifications of regular tree languages, in: V.B. Kudryavtsev, I.G. Rosenberg (Eds.), *Structural Theory of Automata, Semigroups, and Universal Algebra*, Springer, Dordrecht, 2005, pp. 381–432.
- [27] M. Steinby: On the solidity of general varieties of tree languages, *Discuss. Math. Gen. Algebra Appl.* 32 (2012) 23–53.
- [28] M. Takahashi, Generalizations of regular sets and their application to a study of context-free languages, *Inform. Control* 27 (1975) 1–36.
- [29] J.W. Thatcher, Characterizing Derivation Trees of Context-Free Grammars through a Generalization of Finite Automata Theory, *J. Comput. Syst. Sci.* 1(4) (1967) 317–322.
- [30] J.W. Thatcher, Tree automata: an informal survey, in: A.V. Aho (Ed.), *Currents in the Theory of Computing*, Prentice-Hall, Englewood Cliffs, NJ, 1973, pp. 143–172.
- [31] W. Thomas, Logical aspects in the study of tree languages, in: B. Courcelle (Ed.), *9th Colloquium on Trees in Algebra and Programming*, Proc. 9th CAAP, Bordeaux, 1984, Cambridge University Press, London, 1984, pp. 31–49.

## Appendix: Some proofs

This appendix contains several proofs that we have either omitted or just outlined in the main text. Most of them are straightforward, technical and rather uninteresting. In many cases, they can be obtained by obvious modifications from earlier similar proofs.

**Lemma 3.4.** *Let  $\mathcal{A} = (A, \Sigma)$ ,  $\mathcal{B} = (B, \Omega)$  and  $\mathcal{C} = (C, \Gamma)$  be unranked algebras, and  $(\iota, \varphi) : \mathcal{A} \rightarrow \mathcal{B}$  and  $(\varkappa, \psi) : \mathcal{B} \rightarrow \mathcal{C}$  be  $g$ -morphisms.*

- (a) *The product  $(\iota\varkappa, \varphi\psi) : \mathcal{A} \rightarrow \mathcal{C}$  is also a  $g$ -morphism. Moreover, if  $(\iota, \varphi)$  and  $(\varkappa, \psi)$  are  $g$ -epi-,  $g$ -mono- or  $g$ -isomorphisms, then so is  $(\iota\varkappa, \varphi\psi)$ .*
- (b) *If  $R$  is a  $g$ -subalgebra of  $\mathcal{B}$ , then  $R\varphi^{-1}$  is a  $g$ -subalgebra of  $\mathcal{A}$ . In particular, if  $R$  is a  $\Psi$ -subalgebra of  $\mathcal{B}$  for some  $\Psi \subseteq \Omega$ , then  $R\varphi^{-1}$  is a  $\iota^{-1}(\Psi)$ -subalgebra of  $\mathcal{A}$ .*
- (c) *If  $S$  is a  $g$ -subalgebra of  $\mathcal{A}$ , then  $S\varphi$  is a  $g$ -subalgebra of  $\mathcal{B}$ . In particular, if  $S$  is a  $\Psi$ -subalgebra of  $\mathcal{A}$  for some  $\Psi \subseteq \Sigma$ , then  $S\varphi$  is a  $\iota(\Psi)$ -subalgebra of  $\mathcal{B}$ .*

*Proof.* (a) For any  $f \in \Sigma$  and  $w \in A^*$ ,

$$f_{\mathcal{A}}(w)\varphi\psi = \iota(f)_{\mathcal{B}}(w\varphi_*)\psi = (\iota\varkappa)(f)_{\mathcal{C}}((w\varphi_*)\psi_*) = (\iota\varkappa)(f)_{\mathcal{C}}(w(\varphi\psi)_*).$$

Thus  $(\iota\varkappa, \varphi\psi)$  is a  $g$ -morphism from  $\mathcal{A}$  to  $\mathcal{C}$ . Moreover, if  $\iota, \varkappa, \varphi, \psi$  are injective, surjective or bijective, so are also  $\iota\varkappa$  and  $\varphi\psi$ , respectively.

(b) Let  $R$  be a  $\Psi$ -subalgebra of  $\mathcal{B}$  for some  $\Psi \subseteq \Omega$ . To show that  $R\varphi^{-1}$  is a  $\iota^{-1}(\Psi)$ -closed subset of  $\mathcal{A}$ , consider any  $f \in \iota^{-1}(\Psi)$ ,  $m \geq 0$  and  $a_1, \dots, a_m \in R\varphi^{-1}$ . Since  $R$  is  $\Psi$ -closed,  $\iota(f) \in \Psi$  and  $a_1\varphi, \dots, a_m\varphi \in R$ , we get

$$f_{\mathcal{A}}(a_1, \dots, a_m)\varphi = \iota(f)_{\mathcal{B}}(a_1\varphi, \dots, a_m\varphi) \in R,$$

and hence  $f_{\mathcal{A}}(a_1, \dots, a_m) \in R\varphi^{-1}$ .

(c) Let  $S$  be a  $\Psi$ -subalgebra of  $\mathcal{A}$  for some  $\Psi \subseteq \Sigma$ . To see that  $S\varphi$  is a  $\iota(\Psi)$ -closed subset of  $\mathcal{B}$ , consider any  $g \in \iota(\Psi)$ ,  $m \geq 0$ , and  $b_1, \dots, b_m \in S\varphi$ . Then  $g = \iota(f)$  for some  $f \in \Psi$ , and  $b_1 = a_1\varphi, \dots, b_m = a_m\varphi$  for some  $a_1, \dots, a_m \in S$ , and hence

$$g_{\mathcal{B}}(b_1, \dots, b_m) = \iota(f)_{\mathcal{B}}(a_1\varphi, \dots, a_m\varphi) = f_{\mathcal{A}}(a_1, \dots, a_m)\varphi \in S\varphi$$

since  $S$  is  $\Psi$ -closed. □

**Lemma 3.7.** *Let  $\mathcal{A} = (A, \Sigma)$  and  $\mathcal{B} = (B, \Omega)$  be any algebras.*

- (a) *For any  $g$ -congruence  $(\sigma, \theta)$  of  $\mathcal{A}$ , the natural maps  $\theta_{\natural} : A \rightarrow A/\theta, a \mapsto a\theta$ , and  $\sigma_{\natural} : \Sigma \rightarrow \Sigma/\sigma, f \mapsto f\sigma$ , define a  $g$ -epimorphism  $(\sigma_{\natural}, \theta_{\natural}) : \mathcal{A} \rightarrow \mathcal{A}/(\sigma, \theta)$ .*
- (b) *The kernel  $\ker(\iota, \varphi) := (\ker \iota, \ker \varphi)$  of any  $g$ -morphism  $(\iota, \varphi) : \mathcal{A} \rightarrow \mathcal{B}$  is a  $g$ -congruence on  $\mathcal{A}$ . If  $(\iota, \varphi)$  is a  $g$ -epimorphism, then  $\mathcal{A}/\ker(\iota, \varphi) \cong_g \mathcal{B}$ .*

*Proof.* (a) If  $f \in \Sigma$ ,  $m \geq 0$  and  $a_1, \dots, a_m \in A$ , then

$$\begin{aligned} f_{\mathcal{A}}(a_1, \dots, a_m)\theta_{\mathfrak{A}} &= f_{\mathcal{A}}(a_1, \dots, a_m)\theta = (f\sigma)_{\mathcal{A}/(\sigma, \theta)}(a_1\theta, \dots, a_m\theta) \\ &= \sigma_{\mathfrak{A}}(f)_{\mathcal{A}/(\sigma, \theta)}(a_1\theta_{\mathfrak{A}}, \dots, a_m\theta_{\mathfrak{A}}), \end{aligned}$$

which proves  $(\sigma_{\mathfrak{A}}, \theta_{\mathfrak{A}})$  to be a g-morphism. Of course,  $\theta_{\mathfrak{A}}$  and  $\sigma_{\mathfrak{A}}$  are surjective.

(b) Naturally  $\ker \iota \in \text{Eq}(\Sigma)$  and  $\ker \varphi \in \text{Eq}(A)$ , and for any  $f, g \in \Sigma$ ,  $m \geq 0$  and  $a_1, \dots, a_m, b_1, \dots, b_m \in A$ ,

$$\begin{aligned} (f, g) \in \ker \iota, (a_1, b_1), \dots, (a_m, b_m) \in \ker \varphi \\ \Rightarrow \iota(f) = \iota(g), a_1\varphi = b_1\varphi, \dots, a_m\varphi = b_m\varphi \\ \Rightarrow \iota(f)_{\mathcal{B}}(a_1\varphi, \dots, a_m\varphi) = \iota(g)_{\mathcal{B}}(b_1\varphi, \dots, b_m\varphi) \\ \Rightarrow f_{\mathcal{A}}(a_1, \dots, a_m)\varphi = g_{\mathcal{A}}(b_1, \dots, b_m)\varphi \\ \Rightarrow (f_{\mathcal{A}}(a_1, \dots, a_m), g_{\mathcal{A}}(b_1, \dots, b_m)) \in \ker \varphi. \end{aligned}$$

Thus  $\ker(\iota, \varphi)$  is a g-congruence.

Let  $(\iota, \varphi)$  now be a g-epimorphism. We show that  $(\varkappa, \psi) : \mathcal{A}/\ker(\iota, \varphi) \rightarrow \mathcal{B}$  is a g-isomorphism when  $\varkappa : \Sigma/\ker \iota \rightarrow \Omega$  and  $\psi : A/\ker \varphi \rightarrow B$  are defined by  $\varkappa : f \ker \iota \mapsto \iota(f)$  and  $\psi : a \ker \varphi \mapsto a\varphi$ , respectively. First of all, the two mappings are well-defined and injective: for any  $f_1, f_2 \in \Sigma$ ,

$$\varkappa(f_1 \ker \iota) = \varkappa(f_2 \ker \iota) \Leftrightarrow \iota(f_1) = \iota(f_2) \Leftrightarrow f_1 \ker \iota = f_2 \ker \iota,$$

and for any  $a_1, a_2 \in A$ ,

$$(a_1 \ker \varphi)\psi = (a_2 \ker \varphi)\psi \Leftrightarrow a_1\varphi = a_2\varphi \Leftrightarrow a_1 \ker \varphi = a_2 \ker \varphi.$$

Moreover, the maps are surjective because  $\iota$  and  $\varphi$  are surjective. Finally, for any  $f \in \Sigma$ ,  $m \geq 0$ ,  $a_1, \dots, a_m \in A$ ,

$$\begin{aligned} (f \ker \iota)_{\mathcal{A}/\ker(\iota, \varphi)}(a_1 \ker \varphi, \dots, a_m \ker \varphi)\psi \\ = f_{\mathcal{A}}(a_1, \dots, a_m)(\ker \varphi)\psi = f_{\mathcal{A}}(a_1, \dots, a_m)\varphi \\ = \iota(f)_{\mathcal{B}}(a_1\varphi, \dots, a_m\varphi) = \varkappa(f \ker \iota)_{\mathcal{B}}((a_1 \ker \varphi)\psi, \dots, (a_m \ker \varphi)\psi). \end{aligned}$$

Altogether,  $(\varkappa, \psi)$  is a g-isomorphism between  $\mathcal{A}/\ker(\iota, \varphi)$  and  $\mathcal{B}$ .  $\square$

**Proposition 3.17.** *For any  $\Sigma$  and any  $X$ , the term algebra  $\mathcal{T}_{\Sigma}(X)$  is freely generated by  $X$  over the class of all unranked algebras, that is to say,*

- (1)  $\langle X \rangle = T_{\Sigma}(X)$ , and
- (2) if  $\mathcal{A} = (A, \Omega)$  is any unranked algebra, then for any pair of mappings  $\iota : \Sigma \rightarrow \Omega$  and  $\alpha : X \rightarrow A$ , there is a unique g-morphism  $(\iota, (\iota, \alpha)_{\mathcal{A}}) : \mathcal{T}_{\Sigma}(X) \rightarrow \mathcal{A}$  such that  $(\iota, \alpha)_{\mathcal{A}}|_X = \alpha$ .

*Proof.* (1) As  $\langle X \rangle$  is the intersection of those subalgebras of  $\mathcal{T}_{\Sigma}(X)$  that contain  $X$ , it is clear that  $X \subseteq \langle X \rangle \subseteq T_{\Sigma}(X)$ , and that to prove  $T_{\Sigma}(X) \subseteq \langle X \rangle$ , it suffices to show that  $T_{\Sigma}(X) \subseteq B$  for any  $\Sigma$ -closed subset  $B$  of  $\mathcal{T}_{\Sigma}(X)$  for which  $X \subseteq B$ . This can be done by tree induction as follows.

1. The inclusion  $X \subseteq B$  holds by the choice of  $B$ . For any  $f \in \Sigma$ ,  $f = f_{\mathcal{T}_\Sigma(X)}(\varepsilon) \in B$  because  $B$  is  $\Sigma$ -closed. Hence also  $\Sigma \subseteq B$  holds.
2. Let  $t = f(t_1, \dots, t_m)$  for some  $f \in \Sigma$ ,  $m > 0$ , and  $t_1, \dots, t_m \in T_\Sigma(X)$  such that  $t_1, \dots, t_m \in B$ . Then also  $t = f(t_1, \dots, t_m) = f_{\mathcal{T}_\Sigma(X)}(t_1, \dots, t_m) \in B$ .

(2) For any given  $\iota : \Sigma \rightarrow \Omega$  and  $\alpha : X \rightarrow A$ , a g-morphism  $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{A}$  such that  $\varphi|_X = \alpha$  must satisfy the following conditions:

- (1) For  $x \in X$ ,  $x\varphi = \alpha(x)$ .
- (2) For  $f \in \Sigma$ ,  $f\varphi = f_{\mathcal{T}_\Sigma(X)}(\varepsilon)\varphi = \iota(f)\mathcal{A}(\varepsilon)$ .
- (3) For  $t = f(t_1, \dots, t_m)$ , where  $f \in \Sigma$ ,  $m > 0$ ,  $t_1, \dots, t_m \in T_\Sigma(X)$ ,

$$t\varphi = f_{\mathcal{T}_\Sigma(X)}(t_1, \dots, t_m)\varphi = \iota(f)\mathcal{A}(t_1\varphi, \dots, t_m\varphi).$$

Assuming inductively that the values  $t_i\varphi$  are uniquely defined, this defines a unique value for  $t\varphi$ .

It is clear that the thus defined  $(\iota, \varphi)$  is a g-morphism.  $\square$

**Lemma 4.9.** (a)  $S_g S_g = S_g S = S S_g = S_g$ .

$$(b) \ H_g H_g = H_g H = H H_g = H_g.$$

$$(c) \ P_g P_g = P_g P_f = P_f P_g = P_g.$$

$$(d) \ S_g H \leq S_g H_g \leq H S_g \leq H_g S_g.$$

$$(e) \ P_g S \leq P_g S_g \leq S P_g \leq S_g P_g.$$

$$(f) \ P_g H \leq P_g H_g \leq H P_g \leq H_g P_g$$

*Proof.* Statements (a) and (b) hold because obviously  $S_g S_g = S_g$  and  $H_g H_g = H_g$ .

To prove (c), it clearly suffices to show that  $P_g P_g \leq P_g$ . To reduce the notational complexity, we consider g-products with two factors only. In what follows,  $\mathbf{K}$  is always any given class of unranked algebras.

Let  $\mathcal{A}_i = (A_i, \Sigma_i) \in \mathbf{K}$  for  $i \in [4]$  and let  $\tau_1 : \Omega_1 \rightarrow \Sigma_1 \times \Sigma_2$  and  $\tau_2 : \Omega_2 \rightarrow \Sigma_3 \times \Sigma_4$  be mappings, and let  $\mathcal{B}_1 = (B_1, \Omega_1)$  and  $\mathcal{B}_2 = (B_2, \Omega_2)$  be algebras that are isomorphic to  $\tau_1(\mathcal{A}_1, \mathcal{A}_2)$  and  $\tau_2(\mathcal{A}_3, \mathcal{A}_4)$  via the respective isomorphisms  $\varphi_1 : A_1 \times A_2 \rightarrow B_1$  and  $\varphi_2 : A_3 \times A_4 \rightarrow B_2$ . Next, define a mapping  $\lambda : \Gamma \rightarrow \Omega_1 \times \Omega_2$ . Then any algebra  $\mathcal{C} = (C, \Gamma)$  isomorphic to  $\lambda(\mathcal{B}_1, \mathcal{B}_2)$  is a typical member of  $P_g(P_g(\mathbf{K}))$ . We should show that  $\mathcal{C} \in P_g(\mathbf{K})$ .

Define  $\mu : \Gamma \rightarrow \Sigma_1 \times \Sigma_2 \times \Sigma_3 \times \Sigma_4$  so that  $\mu(g) = (f_1, f_2, f_3, f_4)$  if  $\lambda(g) = (h_1, h_2)$ ,  $\tau_1(h_1) = (f_1, f_2)$ , and  $\tau_2(h_2) = (f_3, f_4)$ . Then  $\mu(\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4)$  is a g-product of members of  $\mathbf{K}$ . We show that

$$\psi : \mu(\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4) \rightarrow \lambda(\mathcal{B}_1, \mathcal{B}_2), (a_1, a_2, a_3, a_4) \mapsto ((a_1, a_2)\varphi_1, (a_3, a_4)\varphi_2),$$

is an isomorphism. Clearly,  $\psi$  is a bijection since  $\varphi_1$  and  $\varphi_2$  are bijections. Consider any  $g \in \Gamma$ ,  $m \geq 0$ , and  $(a_{11}, a_{12}, a_{13}, a_{14}), \dots, (a_{m1}, a_{m2}, a_{m3}, a_{m4}) \in A_1 \times A_2 \times A_3 \times A_4$ . If  $\lambda(g) = (h_1, h_2)$ ,  $\tau_1(h_1) = (f_1, f_2)$ , and  $\tau_2(h_2) = (f_3, f_4)$ , then

$$\begin{aligned}
& g_{\mu(\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4)}((a_{11}, a_{12}, a_{13}, a_{14}), \dots, (a_{m1}, a_{m2}, a_{m3}, a_{m4}))\psi \\
&= ((f_1)_{\mathcal{A}_1}(a_{11}, \dots, a_{m1}), (f_2)_{\mathcal{A}_2}(a_{12}, \dots, a_{m2}), \\
&\quad (f_3)_{\mathcal{A}_3}(a_{13}, \dots, a_{m3}), (f_4)_{\mathcal{A}_4}(a_{14}, \dots, a_{m4}))\psi \\
&= (((f_1)_{\mathcal{A}_1}(a_{11}, \dots, a_{m1}), (f_2)_{\mathcal{A}_2}(a_{12}, \dots, a_{m2}))\varphi_1, \\
&\quad ((f_3)_{\mathcal{A}_3}(a_{13}, \dots, a_{m3}), (f_4)_{\mathcal{A}_4}(a_{14}, \dots, a_{m4}))\varphi_2) \\
&= (((h_1)_{\tau_1(\mathcal{A}_1, \mathcal{A}_2)}((a_{11}, a_{12}), \dots, (a_{m1}, a_{m2})))\varphi_1, \\
&\quad ((h_2)_{\tau_2(\mathcal{A}_3, \mathcal{A}_4)}((a_{13}, a_{14}), \dots, (a_{m3}, a_{m4})))\varphi_2) \\
&= ((h_1)_{\mathcal{B}_1}((a_{11}, a_{12})\varphi_1, \dots, (a_{m1}, a_{m2})\varphi_1), \\
&\quad (h_2)_{\mathcal{B}_2}((a_{13}, a_{14})\varphi_2, \dots, (a_{m3}, a_{m4})\varphi_2)) \\
&= g_{\lambda(\mathcal{B}_1, \mathcal{B}_2)}(((a_{11}, a_{12})\varphi_1, (a_{13}, a_{14})\varphi_2), \dots, ((a_{m1}, a_{m2})\varphi_1, (a_{m3}, a_{m4})\varphi_2)) \\
&= g_{\lambda(\mathcal{B}_1, \mathcal{B}_2)}((a_{11}, a_{12}, a_{13}, a_{14})\psi, \dots, (a_{m1}, a_{m2}, a_{m3}, a_{m4})\psi).
\end{aligned}$$

So  $\psi$  is also a morphism, and  $\mathcal{C} \in P_g(\mathbf{K})$  as  $\mathcal{C} \cong \lambda(\mathcal{B}_1, \mathcal{B}_2) \cong \mu(\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4)$ .

In each of (d), (e) and (f), it suffices to prove the second inequality because the first and the third inequalities are obvious.

To complete the proof of (d), we should show that  $S_g H_g \leq H S_g$ . To construct a typical member  $\mathcal{C} = (C, \Gamma)$  of  $S_g H_g(\mathbf{K})$ , let  $\mathcal{A} = (A, \Sigma)$  be in  $\mathbf{K}$ ,  $(\iota, \varphi) : \mathcal{A} \rightarrow \mathcal{A}'$  be a g-epimorphism,  $\mathcal{B} = (B, \Omega)$  be a g-subalgebra of  $\mathcal{A}'$ , and let  $(\varkappa, \psi) : \mathcal{B} \rightarrow \mathcal{C}$  be a g-isomorphism. Now  $\mathcal{B}\varphi^{-1} = (B\varphi^{-1}, \iota^{-1}(\Omega))$  is a g-subalgebra of  $\mathcal{A}$ . If we choose a subset  $\Sigma'$  of  $\iota^{-1}(\Omega)$  so that the restriction of  $\iota$  to  $\Sigma'$  is a bijection  $\iota' : \Sigma' \rightarrow \Omega$ , then  $\mathcal{D} = (B\varphi^{-1}, \Sigma')$  is a g-subalgebra of  $\mathcal{A}$ .

Next we define a  $\Gamma$ -algebra  $\mathcal{E} = (B\varphi^{-1}, \Gamma)$  so that for each  $g \in \Gamma$ ,  $g_{\mathcal{E}} = f_{\mathcal{D}}$  for the  $f \in \Sigma'$  such that  $g = \varkappa(\iota'(f))$ . Then  $(\iota'\varkappa, 1_{B\varphi^{-1}}) : \mathcal{D} \rightarrow \mathcal{E}$  is a g-isomorphism. Indeed, if  $f \in \Sigma'$  and  $w \in (B\varphi^{-1})^*$ , then

$$f_{\mathcal{D}}(w)1_{B\varphi^{-1}} = f_{\mathcal{D}}(w) = \varkappa(\iota'(f))_{\mathcal{E}}(w1_{B\varphi^{-1}}).$$

This means that  $\mathcal{E} \in S_g(\mathbf{K})$ . Next, we show that  $\varphi\psi : \mathcal{E} \rightarrow \mathcal{C}$  is an epimorphism. Clearly,  $B\varphi^{-1}\varphi\psi = \mathcal{C}$ . Consider any  $g \in \Gamma$  and  $w \in (B\varphi^{-1})^*$ . Let  $f \in \Sigma'$  and  $h \in \Omega$  be such that  $\iota'(f) = h$  and  $\varkappa(h) = g$ . Then

$$g_{\mathcal{E}}(w)\varphi\psi = f_{\mathcal{D}}(w)\varphi\psi = f_{\mathcal{A}}(w)\varphi\psi = h_{\mathcal{A}'}(w\varphi)\psi = h_{\mathcal{B}}(w\varphi)\psi = g_{\mathcal{C}}(w\varphi\psi).$$

Thus  $\mathcal{C} \in H S_g(\mathbf{K})$ .

The proof of (e) is complete when we show that  $P_g S_g \leq S P_g$ . Any algebra  $\mathcal{D} = (D, \Gamma)$  in  $P_g S_g(\mathbf{K})$  is isomorphic to a g-product  $\lambda(\mathcal{C}_1, \dots, \mathcal{C}_n)$ , where  $n \geq 0$ , and for each  $i \in [n]$ ,  $\mathcal{A}_i = (A_i, \Sigma_i)$  is a member of  $\mathbf{K}$ ,  $\mathcal{B}_i = (B_i, \Omega_i)$  is a g-subalgebra of  $\mathcal{A}_i$ ,  $\mathcal{C}_i = (C_i, \Gamma_i)$  is an algebra g-isomorphic to  $\mathcal{B}_i$  via some g-isomorphism  $(\iota_i, \varphi_i) : \mathcal{B}_i \rightarrow \mathcal{C}_i$ , and  $\lambda : \Gamma \rightarrow \Gamma_1 \times \dots \times \Gamma_n$  is a mapping. It suffices to show that  $\lambda(\mathcal{C}_1, \dots, \mathcal{C}_n) \in S P_g(\mathbf{K})$ .

To do this, we define the mapping  $\varkappa : \Gamma \rightarrow \Sigma_1 \times \cdots \times \Sigma_n$  so that  $\varkappa(g) = (f_1, \dots, f_n)$  if  $\lambda(g) = (g_1, \dots, g_n)$  and  $\iota_1(f_1) = g_1, \dots, \iota_n(f_n) = g_n$ . Then  $\varkappa(\Gamma) \subseteq \Omega_1 \times \cdots \times \Omega_n$ . Now  $\varkappa(\mathcal{A}_1, \dots, \mathcal{A}_n)$  is in  $P_g(\mathbf{K})$  and  $\varkappa(\mathcal{B}_1, \dots, \mathcal{B}_n)$  is a subalgebra of  $\varkappa(\mathcal{A}_1, \dots, \mathcal{A}_n)$ . Indeed, for any  $g \in \Gamma$ ,  $m \geq 0$  and  $(b_{11}, \dots, b_{1n}), \dots, (b_{m1}, \dots, b_{mn}) \in B_1 \times \cdots \times B_n$ , if  $\varkappa(g) = (f_1, \dots, f_n)$ , then

$$\begin{aligned} g_{\varkappa(\mathcal{A}_1, \dots, \mathcal{A}_n)}((b_{11}, \dots, b_{1n}), \dots, (b_{m1}, \dots, b_{mn})) \\ = ((f_1)_{\mathcal{A}_1}(b_{11}, \dots, b_{m1}), \dots, (f_n)_{\mathcal{A}_n}(b_{1n}, \dots, b_{mn})) \in B_1 \times \cdots \times B_n \end{aligned}$$

since each  $\mathcal{B}_i$  is an  $\Omega_i$ -closed subset of  $\mathcal{A}_i$ . Next we verify that the mapping

$$\varphi : B_1 \times \cdots \times B_n \rightarrow C_1 \times \cdots \times C_n, (b_1, \dots, b_n)\varphi \mapsto (b_1\varphi_1, \dots, b_n\varphi_n),$$

defines an isomorphism from  $\varkappa(\mathcal{B}_1, \dots, \mathcal{B}_n)$  to  $\lambda(\mathcal{C}_1, \dots, \mathcal{C}_n)$ . Clearly,  $\varphi$  is bijective. Moreover, for  $g$ , the  $b_{ij}$ 's and  $f_i$ 's as above,

$$\begin{aligned} g_{\varkappa(\mathcal{B}_1, \dots, \mathcal{B}_n)}((b_{11}, \dots, b_{1n}), \dots, (b_{m1}, \dots, b_{mn}))\varphi \\ = ((f_1)_{\mathcal{B}_1}(b_{11}, \dots, b_{m1}), \dots, (f_n)_{\mathcal{B}_n}(b_{1n}, \dots, b_{mn}))\varphi \\ = ((f_1)_{\mathcal{B}_1}(b_{11}, \dots, b_{m1})\varphi_1, \dots, (f_n)_{\mathcal{B}_n}(b_{1n}, \dots, b_{mn})\varphi_n) \\ = (\iota_1(f_1)_{\mathcal{C}_1}(b_{11}\varphi_1, \dots, b_{m1}\varphi_1), \dots, \iota_n(f_n)_{\mathcal{C}_n}(b_{1n}\varphi_n, \dots, b_{mn}\varphi_n)) \\ = g_{\lambda(\mathcal{C}_1, \dots, \mathcal{C}_n)}((b_{11}\varphi_1, \dots, b_{1n}\varphi_n), \dots, (b_{m1}\varphi_1, \dots, b_{mn}\varphi_n)) \\ = g_{\lambda(\mathcal{C}_1, \dots, \mathcal{C}_n)}((b_{11}, \dots, b_{1n})\varphi, \dots, (b_{m1}, \dots, b_{mn})\varphi). \end{aligned}$$

This shows that  $\lambda(\mathcal{C}_1, \dots, \mathcal{C}_n) \in SP_g(\mathbf{K})$ .

Finally, we prove the critical inequality  $P_g H_g \leq H P_g$  of (f). Let  $n \geq 0$  and for each  $i \in [n]$ , let  $\mathcal{A}_i = (A_i, \Sigma_i)$  be an algebra in  $\mathbf{K}$  and let  $(\iota_i, \varphi_i) : \mathcal{A}_i \rightarrow \mathcal{B}_i$  be a g-epimorphism from  $\mathcal{A}_i$  onto an algebra  $\mathcal{B}_i = (B_i, \Omega_i)$ . Let  $\lambda : \Gamma \rightarrow \Omega_1 \times \cdots \times \Omega_n$  be a mapping and  $\psi : \lambda(\mathcal{B}_1, \dots, \mathcal{B}_n) \rightarrow \mathcal{C}$  an isomorphism from  $\lambda(\mathcal{B}_1, \dots, \mathcal{B}_n)$  to some  $\mathcal{C} = (C, \Gamma)$ . Then  $\mathcal{C}$  is a typical member of  $P_g H_g(\mathbf{K})$ .

Define  $\varkappa : \Gamma \rightarrow \Sigma_1 \times \cdots \times \Sigma_n$  as follows. If  $g \in \Gamma$  and  $\lambda(g) = (h_1, \dots, h_n)$ , choose any  $f_1 \in \Sigma_1, \dots, f_n \in \Sigma_n$  for which  $\iota_1(f_1) = h_1, \dots, \iota_n(f_n) = h_n$  and set  $\varkappa(g) = (f_1, \dots, f_n)$ . Then the mapping

$$\varphi : A_1 \times \cdots \times A_n \rightarrow C, (a_1, \dots, a_n)\varphi \mapsto (a_1\varphi_1, \dots, a_n\varphi_n)\psi,$$

is an epimorphism from  $\varkappa(\mathcal{A}_1, \dots, \mathcal{A}_n)$  to  $\mathcal{C}$ . Firstly,  $\varphi$  is surjective since  $\psi$  and  $\varphi_1, \dots, \varphi_n$  are surjective. Secondly, for any  $g \in \Gamma$ ,  $m \geq 0$ , and  $(a_{11}, \dots, a_{1n}), \dots, (a_{m1}, \dots, a_{mn}) \in A_1 \times \cdots \times A_n$ , if  $\varkappa(g) = (f_1, \dots, f_n)$ , then

$$\begin{aligned} g_{\varkappa(\mathcal{A}_1, \dots, \mathcal{A}_n)}((a_{11}, \dots, a_{1n}), \dots, (a_{m1}, \dots, a_{mn}))\varphi \\ = ((f_1)_{\mathcal{A}_1}(a_{11}, \dots, a_{m1}), \dots, (f_n)_{\mathcal{A}_n}(a_{1n}, \dots, a_{mn}))\varphi \\ = ((f_1)_{\mathcal{A}_1}(a_{11}, \dots, a_{m1})\varphi_1, \dots, (f_n)_{\mathcal{A}_n}(a_{1n}, \dots, a_{mn})\varphi_n)\psi \\ = (\iota_1(f_1)_{\mathcal{B}_1}(a_{11}\varphi_1, \dots, a_{m1}\varphi_1), \dots, \iota_n(f_n)_{\mathcal{B}_n}(a_{1n}\varphi_n, \dots, a_{mn}\varphi_n))\psi \\ = g_{\lambda(\mathcal{B}_1, \dots, \mathcal{B}_n)}((a_{11}\varphi_1, \dots, a_{1n}\varphi_n), \dots, (a_{m1}\varphi_1, \dots, a_{mn}\varphi_n))\psi \\ = g_{\mathcal{C}}((a_{11}\varphi_1, \dots, a_{1n}\varphi_n)\psi, \dots, (a_{m1}\varphi_1, \dots, a_{mn}\varphi_n)\psi) \\ = g_{\mathcal{C}}((a_{11}, \dots, a_{1n})\varphi, \dots, (a_{m1}, \dots, a_{mn})\varphi). \end{aligned}$$

This shows that  $\mathcal{C} \in H P_g(\mathbf{K})$ . □

**Lemma 6.2.** *For any subset  $H \subseteq A$  of an unranked algebra  $\mathcal{A} = (A, \Sigma)$ ,  $\theta_H$  is the greatest congruence of  $\mathcal{A}$  that saturates  $H$ .*

*Proof.* It is clear that  $\theta_H$  is an equivalence on  $A$ . The congruence property follows immediately by Lemma 3.20: if  $a \theta_H b$  and  $q \in \text{Tr}(\mathcal{A})$ , then for any  $p \in \text{Tr}(\mathcal{A})$ ,

$$p(q(a)) \in H \Leftrightarrow p(q)(a) \in H \Leftrightarrow p(q)(b) \in H \Leftrightarrow p(q(b)) \in H,$$

which shows that  $q(a) \theta_H q(b)$ . The congruence  $\theta_H$  also saturates  $H$ . Indeed, if  $a \theta_H b$  and  $a \in H$ , then also  $b \in H$  because  $a = 1_A(a)$  and  $b = 1_A(b)$ .

Finally, let  $\theta$  be a congruence of  $\mathcal{A}$  saturating  $H$ , and consider any  $a, b \in A$  such that  $a \theta b$ . Again by Lemma 3.20,  $p(a) \theta p(b)$  for every  $p \in \text{Tr}(\mathcal{A})$ . Since  $\theta$  saturates  $H$ , this means that, for every  $p \in \text{Tr}(\mathcal{A})$ ,  $p(a) \in H$  iff  $p(b) \in H$ , i.e., that  $a \theta_H b$ .  $\square$

**Proposition 7.5.** *(a)  $\emptyset, T_\Sigma(X) \in \text{Rec}(\Sigma, X)$  and  $\text{Rec}(\Sigma, X)$  is closed under all Boolean operations.*

*Proof.* Clearly,  $\emptyset$  and  $T_\Sigma(X)$  are recognized even by trivial  $\Sigma$ -algebras. If  $T, U \in \text{Rec}(\Sigma, X)$ , then there are regular  $\Sigma$ -algebras  $\mathcal{A} = (A, \Sigma)$  and  $\mathcal{B} = (B, \Sigma)$  such that  $T = F\varphi^{-1}$  and  $U = G\psi^{-1}$  for some morphisms  $\varphi : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{A}$ ,  $\psi : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{B}$  and some  $F \subseteq A$ ,  $G \subseteq B$ . Then  $T_\Sigma(X) \setminus T = (A \setminus F)\varphi^{-1} \in \text{Rec}(\Sigma, X)$  and  $T \cap U = (F \times G)\eta^{-1}$  for the morphism  $\eta : \mathcal{T}_\Sigma(X) \rightarrow \mathcal{A} \times \mathcal{B}, t \mapsto (t\varphi, t\psi)$ . Since  $\mathcal{A} \times \mathcal{B}$  is regular, this means that also  $T \cap U \in \text{Rec}(\Sigma, X)$ .  $\square$